

**Some pages of this thesis may have been removed for copyright restrictions.**

If you have discovered material in Aston Research Explorer which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown policy](#) and contact the service immediately (openaccess@aston.ac.uk)

# Implicit Emotion Detection in Text

Orizu Udochukwu  
Doctor of Philosophy

Aston University  
August 2018

Orizu Udochukwu, 2018 asserts his moral right to be identified as the author of this thesis

*This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without appropriate permission or acknowledgement*

## Aston University

**Title:** Implicit Emotion Detection in Text and Application Areas  
**Student:** Orizu Udochukwu  
**Degree:** PhD  
**Year:** 2018

### Summary

In text, emotion can be expressed explicitly, using emotion-bearing words (e.g. happy, guilty) or implicitly without emotion-bearing words. Existing approaches focus on the detection of explicitly expressed emotion in text. However, there are various ways to express and convey emotions without the use of these emotion-bearing words. For example, given two sentences: “*The outcome of my exam makes me happy*” and “*I passed my exam*”, both sentences express happiness, with the first expressing it explicitly and the other implying it.

In this thesis, we investigate implicit emotion detection in text. We propose a rule-based approach for implicit emotion detection, which can be used without labeled corpora for training. Our results show that our approach outperforms the lexicon matching method consistently and gives competitive performance in comparison to supervised classifiers. Given that emotions such as guilt and admiration which often require the identification of blameworthiness and praiseworthiness, we also propose an approach for the detection of blame and praise in text, using an adapted psychology model, Path model to blame. Lack of benchmarking dataset led us to construct a corpus containing comments of individuals’ emotional experiences annotated as blame, praise or others.

Since implicit emotion detection might be useful for conflict-of-interest (CoI) detection in Wikipedia articles, we built a CoI corpus and explored various features including linguistic and stylometric, presentation, bias and emotion features. Our results show that emotion features are important when using Nave Bayes, but the best performance is obtained with SVM on linguistic and stylometric features only.

Overall, we show that a rule-based approach can be used to detect implicit emotion in the absence of labelled data; it is feasible to adopt the psychology path model to blame for blame/praise detection from text, and implicit emotion detection is beneficial for CoI detection in Wikipedia articles.

Keywords: Implicit emotion detection, OCC model, Blame and praise detection, Conflict-of-Interest detection, Rule-based approaches.

This thesis is dedicated to  
Miss Ifio Akpan, My Family, My Friends  
for their love and support during this past 4 years

## **Acknowledgements**

I would like to thank my supervisors Dr. Yulan He for her amazing patience and support in overcoming numerous obstacles I faced during my research and Dr. David Lowe for his kind suggestions and ideas.

I would like to thank my fellow doctoral students for their feedback, cooperation and of course friendship. Especially, all the members of my reading group for the numerous suggestions and feedback.

I would like to thank my friends for accepting nothing less than excellence from me. Last but not the least; I would like to thank my family: my parents, brother, and sisters for supporting me spiritually throughout writing this thesis and my life in general.

# Contents

<b>1</b>	<b>Introduction</b>	<b>12</b>
1.1	Overview . . . . .	12
1.2	Challenges . . . . .	14
1.3	The Research question, hypotheses, and contributions . . . . .	16
1.4	Overview of Thesis . . . . .	18
<b>2</b>	<b>Literature Review</b>	<b>19</b>
2.1	Emotion Analysis . . . . .	19
2.1.1	Theories of Emotion . . . . .	19
2.1.2	Classification of Emotion . . . . .	20
2.1.3	Computational Models for Emotion . . . . .	22
2.1.3.1	The OCC Model . . . . .	22
2.1.4	Emotion Detection Approaches . . . . .	31
2.1.4.1	Lexicon-based Approaches . . . . .	31
2.1.4.2	Learning-based Approaches . . . . .	32
2.1.4.3	Supervised Learning . . . . .	34
2.1.4.4	Semi-supervised Learning . . . . .	36
2.1.4.5	Unsupervised Learning . . . . .	36
2.1.4.6	Discussion . . . . .	37
2.1.4.7	Hybrid Approaches . . . . .	38
2.1.4.8	Other Detection Approaches . . . . .	38
2.1.5	Implicit Emotion Detection . . . . .	39
2.1.6	Summary on Emotion Detection . . . . .	40
2.2	Blame and Praise Detection . . . . .	41
2.2.1	Theories of Blame . . . . .	42
2.2.1.1	Cognitive Theories of Blame . . . . .	43
2.2.1.2	Emotional Theories of Blame . . . . .	43
2.2.1.3	Functional Theories of Blame . . . . .	44
2.2.1.4	Conative Theories of Blame . . . . .	44

2.2.2	Blameworthiness . . . . .	44
2.2.3	Path Model of Blame . . . . .	45
2.2.4	Summary on Blame and Praise Detection . . . . .	47
2.3	Conflict of Interest Detection . . . . .	47
2.3.1	Bias . . . . .	50
2.3.2	Vandalism . . . . .	51
2.3.3	Summary on Conflict of Interest Detection . . . . .	52
2.4	Summary . . . . .	52
<b>3</b>	<b>A Rule-Based Approach to Implicit Emotion Detection in Text</b>	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Our Approach . . . . .	54
3.2.1	Overall Steps . . . . .	54
3.2.2	Polarity Detection . . . . .	58
3.2.2.1	Keyword Detection . . . . .	58
3.2.2.2	Negation Handling . . . . .	58
3.2.2.3	Polarity/Sentiment Detection based on Majority Voting . . . . .	61
3.2.3	Assigning Values to the OCC Variables . . . . .	66
3.3	Experiments . . . . .	70
3.3.1	Results and Analysis . . . . .	72
3.3.2	Error Analysis . . . . .	74
3.3.3	Discussion . . . . .	78
3.4	Improving Rule-Based Approach to Implicit Emotion Detection in Text . . . . .	79
3.4.1	Tense Detection . . . . .	79
3.4.1.1	Present Tense . . . . .	81
3.4.1.2	Past Tense . . . . .	81
3.4.1.3	Future Tense . . . . .	83
3.4.2	Using Word Relatedness Measure to Improve Implicit Emotion Detection . . . . .	85
3.4.2.1	Extracting Emotion-bearing Words . . . . .	85
3.4.2.2	Emotion Vector Representation . . . . .	87
3.4.2.3	Sentence Level Analysis . . . . .	89
3.4.2.4	Experiment Results . . . . .	89

<b>4</b>	<b>Detection of Expressions of Blame and Praise in Text</b>	<b>92</b>
4.1	Introduction . . . . .	92
4.2	Our Approach . . . . .	93
4.2.1	Event Detection . . . . .	95
4.2.2	Agent Causality . . . . .	96
4.2.3	Foreseeability . . . . .	97
4.2.4	Coercion . . . . .	99
4.3	Corpus Creation . . . . .	100
4.4	Experiments and Analysis . . . . .	102
4.4.1	Error Analysis . . . . .	103
<b>5</b>	<b>Content-Based Conflict of Interest Detection for Articles in Wikipedia</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Our Approach . . . . .	110
5.2.1	Stylometric Features . . . . .	110
5.2.2	Bias Features . . . . .	111
5.2.3	Presentational Features . . . . .	111
5.2.4	Emotion Features . . . . .	112
5.3	Experiments . . . . .	113
5.3.1	Data . . . . .	113
5.3.2	Preprocessing . . . . .	114
5.3.3	Feature Selection . . . . .	115
5.3.4	Binary Classification Results . . . . .	116
5.3.5	One-Class Classification . . . . .	118
5.3.6	Comparison with an Existing Approach to Vandalism Detection . . . . .	119
5.3.7	Discussion . . . . .	120
<b>6</b>	<b>Conclusion and Future Work</b>	<b>122</b>
6.1	Analysis and Evaluation . . . . .	123
6.2	Contributions . . . . .	124
6.3	Future work . . . . .	125
<b>A</b>	<b>List of Publications</b>	<b>127</b>
<b>B</b>	<b>Annotation Sample and Guidelines</b>	<b>128</b>
<b>C</b>	<b>List of Intensifiers</b>	<b>130</b>





## List of Figures

2.1	Plutchiks Wheel of Emotions [263]. . . . .	21
2.2	The original OCC model [191]. . . . .	24
2.3	Combination for Compound emotions [191]. . . . .	26
2.4	The revised OCC model [246]. . . . .	28
2.5	The PAD model [127]. . . . .	30
2.6	The Path Model of Blame [145]. . . . .	46
3.1	Dependency parser output sample. . . . .	57
3.2	Polarity/Sentiment Detection based on Majority Voting . . . . .	63
3.3	Multi-emotion detection distribution using SemEval dataset . . . . .	73
3.4	List of auxiliary verbs and their base forms [274] . . . . .	80
3.5	List of modal auxiliaries [274] . . . . .	80
3.6	Present tense auxiliary examples [36] . . . . .	82
3.7	Past tense auxiliary examples [36] . . . . .	83
3.8	Future tense auxiliary examples [36] . . . . .	83
3.9	Our Approach . . . . .	86
3.10	Fallback Flow chart . . . . .	86
4.1	Revised Path Model of Blame. . . . .	94
4.2	An example dependency parse result. . . . .	96
4.3	List of lexicographer files with file numbers and a brief description of file's contents. . . . .	98
4.4	Annotation User Interface. . . . .	101
5.1	Two sample documents with features identified by our approach. Words/phrases underlined in text are those which be found in an emotion or sentiment lexicon. Due to space constraint, we only show some key features here such as the emotion/sentiment label, bias score, praise/blame indicator, and sentence type. . . . .	109
5.2	The overall work flow of our proposed approach. . . . .	111
B.1	User Interface sample. . . . .	128

## List of Tables

2.1	OCC Model emotions and definitions [191]	25
3.1	Rules for emotion detection.	55
3.2	Tense Patterns	58
3.3	Variable assignment and possible values.	69
3.4	Statistics of the three datasets used in our experiments. “Total” denotes the original number of sentences in each emotion category while “Implicit” denotes the number of sentences which do not contain any emotion words according to WordNet-Affect.	72
3.5	Performance comparison of F-measure (F) results on the 3 datasets. Alm dataset(F1), ISEAR(F2) and SemEval(F3).	73
3.6	Detection examples of correct and failed detection using our approach on SemEval dataset	75
3.7	Detection examples of correct and failed detection using our approach on ISEAR dataset	75
3.8	Detection examples of correct and failed detection using our approach on Alm dataset	78
3.9	Present tense Rules	82
3.10	Past Tense Rules	82
3.11	Future Tense Rules	84
3.12	Performance comparison of F-measure (F) results on the 3 datasets. Alm dataset(F1), ISEAR(F2) and SemEval(F3).	85
3.13	Top 5 related words returned for <i>shoplifting</i> before and after retrofitting	89
3.14	Tables showing results of the ALM data set	90
3.15	Tables showing results of the ISEAR data set	90
4.1	Annotation agreement matrix.	101
4.2	Classification results of <i>blame, praise or others</i> .	102
4.3	Classification results of <i>self-blame, self-praise, blame-others, praise-others</i> or <i>others</i> .	103
4.4	Detection examples of correct and failed detection using our approach on the dataset	104
5.1	Stylometric features.	112

5.2	Bias features and subtypes. . . . .	113
5.3	Definitions of 10 presentational relations used (N stands for nucleus, R for reader and W for writer). . . . .	113
5.4	Individual feature ranking results from Correlation-based Feature Subset Selection (CFS) and Information Gain Ratio (IGR). . . . .	116
5.5	Merged features from the feature selection results from Correlation-based Feature Subset Selection (CFS) and Information Gain Ratio (IGR). . . . .	116
5.6	Conflict-of-Interest (CoI) detection results in Precision, Recall and F-measure using SVM, MaxEnt and NB with various feature sets. . . . .	117
5.7	CoI detection results using one-class classification. . . . .	118
5.8	Comparison with an existing approach to vandalism detection. . . . .	119

## **Chapter 1**

### **Introduction**

#### **1.1 Overview**

The internet holds huge document collections which can provide various types of information [180, 289]. Studies in various areas such as Natural Language Processing (NLP), Affective Computing and Computational Linguistics have begun to conduct various researches on these text document collections [180, 63]. This text not only contains information that describes facts and events, they can in most cases hold information about the writer's attitudes or stances towards the subject matter [75, 135, 62, 199].

The emergence of the web, combined with various social media platforms have changed the way people live, with a constantly growing number of people sharing thoughts, opinions, and emotions in these environments. Analyzing these contents from all kinds of platforms can provide both public and private sectors with opportunities to grow and understand the general public. Private sectors can use sentiment analysis results to track the view of the public on their services/products, as well as the overall image of their organizations, as opinions published and shared on these online platforms can have an effect on the reputation of the organization [295, 282]. These organizations have realized that understanding public opinions from these platforms can help build good relationships with their customers and respond appropriately to market and socio-economic changes. For the public sector, various studies have shown the close relationship between events on social media platforms and various issues in politics, for example, the role social media platforms played during civil uprising in various Arab countries (known as Arab Spring). There was a ten-fold increase in the number of tweets talking about a need for Egypt to have a political change, resulting in the resignation of the

Egyptian president [35, 104].

There is evidence that human emotions influence the design of software products, evidence of how products also can influence human feelings and evidence of how those feelings can make people buy or not. The points above show how important it is to detect and identify sentiments and emotions in text. Key influences of our behavior as humans are our beliefs, view of reality and the choices we make. It is constantly becoming difficult even in journalism to find objective information because these beliefs make one write more subjective sentences than objective ones [122, 101]. Subjective sentences are sentences which express personal views, beliefs, and feelings while objective sentences present some factual information [140].

The concept of emotion is a complex one and the term is often used in various ways. An emotion is considered a short-lived episode of coordinated brain activity, which creates specific changes in response to both external and internal events [75, 135]. The subjective experience of these emotions is often referred to as feelings (to feel fear, joy, or guilt). Feelings are subjective and best described as the way an individual experiences an emotion [75]. Moods are a more spread out affective state, with smaller intensity when compared to emotion but are of a more prolonged in duration than emotion [75]. For example: receiving an “A” in one’s exam might put one in a good mood all day, however, the emotion of happiness is short-lived and very intense [75].

Natural Language Processing (NLP) areas such as subjectivity and sentiment analysis have more developed foundational methodologies which are beneficial to emotion analysis [112, 116, 166]. Emotion analysis in text aims at identifying and extracting emotions towards a specific subject matter. Given its importance, much research in the area of emotion recognition in text has been ongoing in recent years and the attention the area is getting is constantly growing. Previous research on emotion analysis of text focused on text content from web-blogs, stories, news, text messages, spoken dialogs, etc [140, 235, 112, 199]. Emotion has been studied in depth in the field of psychology and behavioural sciences, mainly as a result of its importance to the human nature and this has led to the growth in this area of computing [251, 33, 52, 82, 93, 233].

Emotions play a big role in human intelligence, decision making, creativity, social interaction and more [135, 75, 62, 61]. The study of emotions is very broad and includes areas such as emotional responses to physiological reactions, gestures, and postures based on subjective experiences, facial expressions and emotion intensities. There is a strong relationship between emotions and

sentiments. The intensity of an emotion can typically be linked to the strength of a sentiment or opinion it produces.

Emotion is subjective and expressed with the use of subjective language, a language used mainly for expressing opinions and/or evaluations [140, 135]. It can hence be said that emotion analysis in text benefits from the ability to detect and distinguish objective from subjective language. Clues on the attitude/position of the writer on a subject matter, are often provided in the lexical choices of the writer, also in the structure of the text. These clues can be as a result of emotions, their writing style or the presentation relations used in the text. These clues can be identified using a combination of various NLP techniques and research from areas such as emotion and sentiment analysis in text, detection of bias and many more. As previously mentioned, distinguishing objective from subjective language is important and recent studies have shown that both languages when combined is effective for accurate results when compared to methods that rely solely on subjective language as shown in the area of subjectivity analysis [140].

In this thesis, we investigate emotion analysis, aiming mainly at the implicit emotion detection problem, which deals with classifying text documents into emotion categories in the absence of words considered as emotion bearing words. In particular, we research several methods for extracting emotion and sentiment in text. Our research in implicit emotion detection leads us to identify that certain emotion categories such as "*Pride*", "*Shame*", "*Gratitude*", "*Admiration*", "*Remorse*", and many more require one to identify when the subject of the emotion is praiseworthy or blameworthy. This leads us to study and investigate the detection of expressions of blame and praise in text documents. Therefore, we propose adopting the psychology "*Path Model for Blame*" into a model that can be applied to text documents. We use various datasets in the experimental work and annotate a new dataset for the detection of blame. We study various applications of implicit emotion analysis, looking at document types that should have less explicit expressions of emotions as in reference sites like Wikipedia.

## 1.2 Challenges

As previously stated, emotion is subjective and with subjectivity, only the bearer (writer) truly knows the underlining motivation. Hence, there are a few challenges to be tackled in emotion detection from text such as:

- Natural language is ambiguous, that is; words and sentences can have different meanings [106] “The meaning of a word is its use in the language”. For example, “*Pride*” by definition: “is a feeling that one respects oneself and deserving of respect by other people” or “a feeling that you are more important or better than other people”. In both cases, the word is spelt the same way but these are clearly different [106, 259]. A sentence can also be ambiguous even if the words which make up the sentence are not. One type of such ambiguity is structural ambiguity. For example: “*Stolen rifle found by tree*”.
- The lack of labeled data required for classifiers training [112, 199].
- Identify emotions in text when no emotion keywords or phrases have been used. This presents the problem of trying to identify opinionated views in a document when the writers are avoiding the use of explicit opinionated words [235, 199].
- Detecting emotion based on only the text content of the article without additional context. For example in the sentence “*When I was so sure I failed an exam but I did not fail*” expresses joy but none of the words are positive words. Additional context (“passing an exam”) is needed to correctly detect the emotion from the aforementioned example correctly [140, 112, 199].
- A single compound/complex sentence can contain multiple subjects and objects, hence it can express multiple emotions and views, which can be directed towards various subjects/objects within a single sentence [140, 235, 112, 199].

Above are some of the major challenges when working on emotion detection in text. However, due to time constraints, we would not be able to focus completely on all 5 points. In this thesis, we try to address the following 3 challenges.

1. **Identify emotions in text when no emotion keywords or phrases have been used:** This is the primary focus of our work and we propose a rule-based emotion detection approach for implicit emotion detection.
2. **The lack of labeled data required for classifiers training:** To help mitigate this challenge, we propose an approach that is independent of labeled data for emotion detection. We also



provide an annotated dataset for the detection of blame and praise, which are useful variables for the detection of emotions like guilt, admiration and many more.

3. **Natural language word level ambiguity:** We try to address this challenge by incorporating contextual valence detection and context level word sense disambiguation. These techniques help us to understand the context meaning of the words within the sentences they appear in. Identifying when an entity in an article subject matter is overly praised/blamed, the style of writings found in the text can help shed more light on the intentions of the writer.

We also touch briefly the detection of multiple emotions in a sentence but will not go into an in-dept analysis of what these may entail.

In the rest of this chapter, we present our research questions which we will be addressing and the hypotheses for our research. Next, we outline our contributions and a layout for the overall thesis.

### **1.3 The Research question, hypotheses, and contributions**

The main research question investigated in this thesis is:

*How can we detect and correctly classify text documents into emotion categories in the absence of emotion bearing words and labeled corpora?*

The primary focus of this research is to detect implicit emotion in text. We hypothesize that by adapting existing psychology models for use in text analysis, we are able to detect implicit emotions from text without the use of labeled corpora.

To solve/answer the main question, the following sub-questions need to be addressed:

1. *Can we identify implicit expressions of emotion in text without the use of labeled data ?*

There is evidence which shows that an author's emotions towards a subject matter can affect his/her writing on the subject. Emotion detection in text is an area that much work has been done in recent years. However, most existing studies either rely on keyword matching based on emotion lexicons or supervised classifier training from labeled data. We instead propose a rule-based approach for implicit emotion detection, using a model proposed by Ortony, Clore and Collins in their book "The Cognitive Structure of Emotions" often referred to as

the OCC-Model (OCC is taken from the names of the book authors) [191].

2. *Can we detect praise and blame in text, elements necessary for the detection of emotions such as guilt, remorse and many more ?*

Emotions such as guilt require one to first establish oneself as “blameworthy” before the emotion can be said to exist. We propose an approach for blame/praise detection based on the “Path Model of Blame” presented by Malle et al [145].

3. *Can the intelligent combination of emotion features help with conflict-of-interest detection in Wikipedia articles ?*

The conflict-of-interest problem is a complex one and we believe that emotions features as well as other types of features such as linguistic, stylometric and bias features can aid in the classification of such documents. We create a dataset from Wikipedia, investigate various types of features and various feature combinations for the classification of conflict-of-interest articles.

Our contributions from this research can be summarised as follows:

- We propose a new rule-based approach for implicit emotion detection using the OCC model, which detects expressions of emotion in sentences and investigate the existence of multiple expressions of emotions within a single sentence, which is largely ignored in existing work.
- We extend and annotate a dataset of over 7000 comments for blame and praise detection. We propose an approach adapted from Path Model of Blame for detecting expressions of blame and praise in text
- We conduct experiments and take into account the direction of emotions and blame (to see if they are directed towards self or others), something we believe most existing work did not take into consideration.
- We investigate the application of implicit emotion detection on conflict-of-interest (CoI) detection in Wikipedia articles. In specific, we create a dataset and explore various features apart from emotions for this task.

We believe that our work in this thesis is the first to introduce and investigate implicit emotion detection using the OCC model in text documents, blame and praise detection in text and conflict of interest detection in Wikipedia articles based solely on content.

## **1.4 Overview of Thesis**

The structure of the rest of this thesis is as follows:

In **Chapter 2**, we conduct a literature review of the emotion detection task and popular approaches. We also review recent work on the related tasks such as vandalism detection and bias detection.

In **Chapter 3**, we propose a rule-based approach for implicit emotion detection. We also explore multiple emotion detection from a single sentence and evaluate our work on three different datasets. This chapter addresses the first research sub-question.

In **Chapter 4**, we describe our work on detecting expressions of blame in text, presenting a model for blame detection in text. This chapter addresses the second research sub-question.

In **Chapter 5**, we present our work on conflict of interest detection using emotion, presentation, stylometric and linguistic features. Our third research sub-question is addressed in this chapter.

In **Chapter 6**, we conclude our work in this thesis and outline our contributions and future work.

## **Chapter 2**

# **Literature Review**

### **Overview**

In this chapter, an overview of the emotion analysis literature is provided. In particular, the emotion analysis problem is first introduced in Section 2.1. A literature review of the emotion analysis research including a discussion of various approaches, their main strengths, limitations, and gaps are provided in Section 2.2. Recent work on implicit emotion detection is discussed in Section 2.3. Finally, a review of bias and vandalism detection is given in Section 2.4.

### **2.1 Emotion Analysis**

There has been much research in sentiment analysis and emotion analysis of text [81, 76, 285]. This section presents the literature of the previous work in emotion analysis, identifies their strengths and weaknesses, and finally gives an overall comparison.

#### **2.1.1 Theories of Emotion**

Since early studies on emotion in the field of psychology, there have been controversies on what actually causes emotions. There are many theories of emotion. Here are the three most frequently mentioned in research from which all other theories originate.

In the James-Lange theory, the event which causes arousal and physical changes is interpreted as the emotion [75]. For example: if you are taking a walk in the city at night and a man jumps in front of you holding what appears to be a gun, according to James-Langes view, seeing the man would cause an increase in your heart rate; you become aware of your heart beating faster and you experience fear.

In the Cannon-Bard theory of emotion, emotion can be triggered by an event itself [135]. This

theory states that the arousal caused by an event is not necessary for emotion. Thus, an event creates both the emotion and physiological changes, without a dependency on one another. The cognitive theory of emotion is similar to James-Langs theory, it states that emotions have more to do with the interpretation process. Thus, it is the complete scenario and not just the arousal and physical changes that determine emotion [75, 135].

Hot cognition is basically cognition coloured by feeling [39]. There are theories that state that almost all human activities tend to implicate emotions in one manner or the other. This means that day to day events are in some way controlled by our feelings and the expression options available to us [62]. Our choice of words will tend to communicate information about our opinions, beliefs, and judgments [76, 135, 62]. According to Brand, using a stage model one can show how cognition and affect interact: “*before, during, and after*” writing. The “Before” affects are enthusiastic [39]. The “During” affects enables the continuation of writing towards its closure. The “After” affects are the outcome emotions that lead to the next writing episode [39]. Emotions influence both what and how we write, emotions also influence our writing process and our thinking process. Human personal experience tells us that when one is feeling angry and when one is feeling happy lead to different emotional outcomes and actions. A similar thing happens in writing, feeling angry and feeling happy lead to different writing events [39].

Alice in her work [39] stated that the writing psychology must include cognitive as well as affective attributes, because writing which requires thinking and analysis. This representation of the human thought process employs both inductive and deductive reasoning. During writing, the act of thinking deals with words and the sense of the word meaning in context. This sense is a collection of associated psychological events that relate to the word. These word senses are filled with affect [40].

### **2.1.2 Classification of Emotion**

In the study of human emotions, there are various classifications which have been presented over the years. Various researchers have postulated different sets of human emotions, Aristotle maintained that there were 15 basic emotions; Descartes listed 6 emotions [135, 263].

In the Book “Emotions In Social Psychology” [198], the author classified emotions and placed *Love, Joy, Anger, Sadness, Fear, and Surprise* on the primary level and explained the emotion sys-

tem. Ekman made a case for these basic emotions, stating that each emotion has unique and shared characteristics that distinguish them [68].

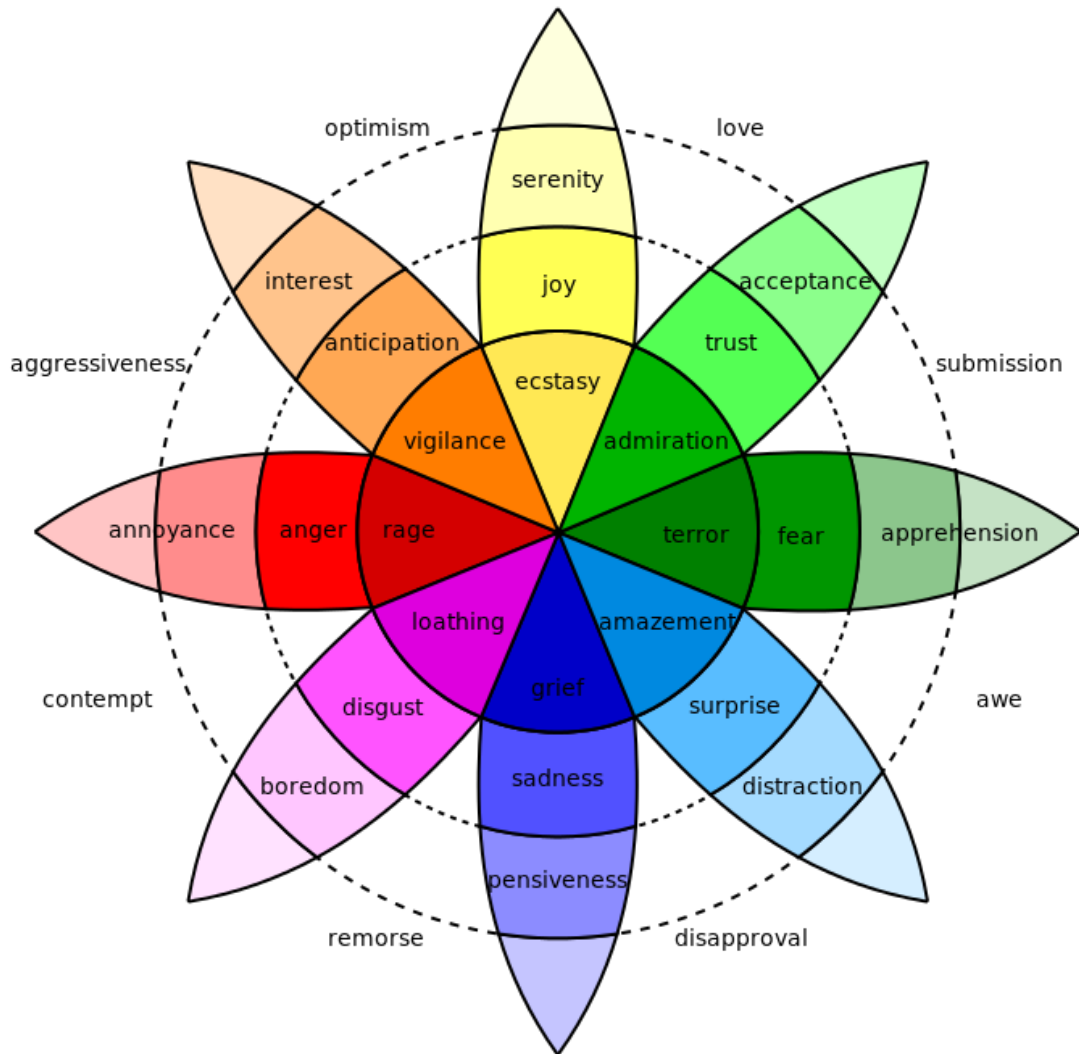


Figure 2.1: Plutchik's Wheel of Emotions [263].

Robert Plutchik created the “wheel of emotions” shown in Figure 2.1. He showed how various emotions can combine with each other to create new emotions. Starting with 8 primary emotions: trust versus disgust; surprise versus anticipation; anger versus fear and joy versus sadness, more advanced emotions were identified based on differences in intensity [263]. Parrot presented a theory, where he identified over 100+ emotions. He then used a tree structure to conceptualize them [198].

In computational linguistics, it is becoming important to be able to detect human emotions and this has led to various computational models of emotion [199].

### **2.1.3 Computational Models for Emotion**

In this section, we look at the various computational models for emotion and try to reveal their uses, techniques, and assumptions. Emotions are a big part of our day to day life as humans, and as a result, there are commonsense interpretations to emotion research terminologies (appraisal, emotion, feeling), which can differ from the technical definition of the said terms within a given computational model for emotion [154].

#### **Appraisal Theories**

Appraisal theory is a predominant theory on emotion and a good source for parties interested in AI systems. Appraisal theory explains the link between cognition and emotion [154]. According to this theory, emotions stem from the types of judgment one makes about an event and one's beliefs, intentions as well as desires. These patterns are called the person-environment relationship [129]. These judgments, are technically referred to as appraisal variables [154]. They characterize parts of the significance of an event to an individual. This theory presents appraisal as central to emotion process. However, various appraisal theorists view "emotion" as an unattached component, while others refer to appraisal configurations such as physical changes and subjective experience as emotion. There is research work that focuses on the interdependence between appraisal variables and emotion labels [191]. In the Book "The Cognitive Structure of Emotion" Ortony, Clore, and Collins defined emotions as valenced reactions, stating that how an individual understands the eliciting conditions determines the emotion.

#### **2.1.3.1 The OCC Model**

The OCC model is a model under the appraisal theory of emotion. The model starts off by providing a few assumptions on three major focus points when dealing with emotion types which include: *Events* which people focus on when they are interested in the consequence of an event. *Agents* which one focuses on because of their actions and *Objects* which one focuses on when one is interested in the properties or aspects. The OCC Model provides a clear and convincing structure of the eliciting conditions of emotions and the variables that affect emotional intensities [191]. It

describes a hierarchy that classifies 22 emotion types. The hierarchy contains three branches. The first branch is the “*consequences of events*” branch. The emotions on this branch express pleasure or displeasure with event consequences. The second branch contains emotions in relation to “*actions of agents*”. An agent can be self or other and these are the attribution type of emotions. When well-being and attribution type emotions are mixed, we get the compound emotions such as anger and remorse. The third branch contains emotions relating to liking or disliking in regards to “*aspects of objects*”. The authors explain how focusing on events, agents, and objects can lead to various emotion. Take an individual finding out that his neighbor is a child-beater. If one focuses on the neighbor as an *agent* of beating, one will judge it as blameworthy as it violates standards. The resulting emotional reaction towards the neighbor will be one of attribution such as reproach. If one focuses on the undesirable child-beating event, it will cause him/her distress. If one focuses on the children, one will feel pity for them. Thus by focusing on different parts of the same situation, one can experience various emotions.





Figure 2.2: The original OCC model [191].

Table 2.1: OCC Model emotions and definitions [191]

<b>Emotion</b>	<b>OCC Model Definition</b>
<b>Well Being Emotions</b>	
Joy	pleased about a desirable event
Distress	displeased about an undesirable event
<b>Prospect Emotions</b>	
Hope	pleased about the prospect of a desirable event
Fear	displeased about the prospect of an undesirable event
Satisfaction	pleased about the confirmation of the prospect of a desirable event
Fear confirmed	displeased about the confirmation of the prospect of an undesirable event
Relief	pleased about the disconfirmation of the prospect of an undesirable event
Disappointment	displeased about the disconfirmation of the prospect of a desirable event
<b>Fortune of Others</b>	
Happy-for	pleased about an event desirable to someone else
Resentment	displeased about an event desirable to someone else
Gloating	pleased about an event undesirable to someone else
Sorry-for	displeased about an event undesirable to someone else
<b>Attribution Emotions</b>	
Pride	approving of one's own praiseworthy action
Shame	disapproving of one's own blameworthy action
Admiration	approving of someone else's praiseworthy action
Reproach	disapproving of someone else's blameworthy action
<b>Compound Emotions</b>	
Gratitude	approving of someone's praiseworthy action and pleased about the related desirable event
Anger	disapproving of someone's blameworthy action and displeased about the related undesirable event
Remorse	disapproving of someone's blameworthy action and displeased about the related undesirable event
Gratification	approving of someone's praiseworthy action and pleased about the related desirable event
<b>Attraction Emotions</b>	
Love	liking an appealing object
Hate	disliking an unappealing object

Looking at the branch which reached from CONSEQUENCES OF EVENTS this branch connects to being pleased and displeased. Here, being pleased or displeased refers to unspecific emotion state that are valenced reactions to an event. This branch continues to reflect if the person experiencing the emotion is reacting to an event focusing on himself or to an event focusing on another person. This gives the two branches CONSEQUENCES FOR SELF and CONSEQUENCES FOR OTHER. Events can be in the past, future or concurrent(present) and can in turn, lead to different types of emotion. The CONSEQUENCES FOR SELF the first group of emotions to consider here are emotions which deal with the well being of self. These emotions do not take the prospect of an events consequence into consideration but are simply what it is like to be pleased or displeased with an event that affects one directly. Joy and distress are two emotions that correspond to one focusing only on the desirability and undesirability of an event. Continuing under the CONSEQUENCES FOR SELF branch we look at the prospect of emotions, these emotions are reactions to the prospect of an event. It involves looking at events that are yet to occur, already occurred or failed to occur.

With prospect-emotions, they are separated by the *status* of the event in question. First, emotions where the experiencer of the emotion is unaware of if the event has occurred or has not occurred, is referred to as an *unconfirmed* status. Second, the individual is aware that the event has occurred, this is a *confirmed* status. Finally, the individual is aware that the event failed to occur, this is a *disconfirmed* status. Looking at the CONSEQUENCES FOR OTHER branch, the event emotions here are in relation to what happens to other people. For example, if one is pleased for someone who is having a good fortune, the basis of this could just be that they like the person and wish them well. The branch with ACTIONS OF AGENTS are for attribution emotions, as they deal with emotions of responsibility for actions. A central part of this is our ability to find a person blameworthy/praiseworthy for an action. With attribution emotion, the primary focus is the agent and his/her causal role rather than the outcome. The model also discusses compound emotions, which are derived by conditions of two different emotions. These emotions generally result from the combination of well-being emotions and attribution emotions as presented in figure 2.3.



Figure 2.3: Combination for Compound emotions [191].

The model also looks at attraction emotions, which are object-based emotions. The appeal- ingness of an object connected to how one associates/categorises objects and one's disposition to objects in said categories and said object's attributes. The model also makes it clear that familiarity has a correlation with liking or not liking an object. According to the model, a one's evaluation of a given emotion-inducing event is based on three variables *desirability*, *praiseworthiness* and *ap-*

*pealingness*, which respectively apply to **event-based**, **agent-based**, and **object-based emotions**. It then goes into detail to identify what it called *global* and *local* variables which affect the intensity of the emotion. The global variables are variables that affect all emotions, while local variables are those that affect only specific groups of emotions. The model is dependent on its variables and the rules for implementing/identifying emotion. The variables are grouped into *emotion-inducing variable* and *emotion intensity variables*. For a full list of the emotions and variables, see [191]. Depending on what is expressed, some of the variables may or may not have assigned values. Emotion recognition aims to infer emotions from text by applying a set of pre-defined rules. Multiple emotions can be inferred from a given situation.

The OCC model is rather complex and full of ambiguity. Steunebrink et al [246], outlined a number of issues of the original OCC model and proposed changes to remove duplications and ambiguities in the original OCC model. The result of these resolutions is evident as shown in the revisited OCC model in Figure 2.4. In our work, we use the revisited OCC model [246] for emotion detection in text.



Figure 2.4: The revised OCC model [246].

Computational appraisal models create complex mechanisms for detecting and extracting appraisal variables. Emotion in these models is not modeled as elaborately as the appraisal variables, mostly treated as a label and in most cases without reference to intensity levels [154]. The modeling of appraisal is such that it is the trigger that results in emotion, derived via rules placed on appraisal variables [154].

The Appraisal theories of emotion are the most dominant emotion theories primarily because it explains and emphasises the connection in relation to the process of emotion appraisal. Appraisal theories have been used with great success in Artificial Intelligence systems and other machine learning systems. The values of this theory to the understanding of emotion is one of the major

reasons we choose to work with a model from the Appraisal theories.

### **Dimensional Theories**

Dimensional theories state that emotions are not unattached entities, but points in a measurable space [154]. Most computational dimensional models stem from the three-dimensional “PAD” model presented by Mehrabian et al [154], which includes a measure of valence called *pleasure*, affective activation known as *arousal*, and a measure of control called *dominance*. Dimensional theory focuses on different emotion components and their relationships in comparison to appraisal theories [154]. The model focuses more on core affect/moods and less on emotions. To the dimensional model, core affect is a non-rational construct of a non-intentional state, which provides a summarized state of the individual [127]. Core affect is presented as a continuous time varying process that is represented at a given period in time by a 3D space controlled by events. Computational models which stem from these theories, dwell on the processes that are related to affect as it is affected by surrounding events [154].



Figure 2.5: The PAD model [127].

### **Anatomic Approaches**

These try to re-create an organism's neural links for emotional reaction. They focus on sub-symbolic processes and view emotions as discrete neural circuits. They then focus on the processes which are associated with these discrete neural circuits. Thus, models inspired by these theories often rely on process assumptions, and not as comprehensive as appraisal or dimensional theories [154]. Researchers who work with such models often focus on a specific emotion. Anatomic models often present a dual perspective of emotion, "a fast, automatic, undifferentiated emotional response" and "a slower, differentiated response that relies on higher-level reasoning processes" [154].

## **Rational Approaches**

These approaches view emotion for its adaptive function, and this function is abstracted and incorporated into a model of intelligence [154]. Models that stem from these theories are directed towards improving theories of machine intelligence [154]. They focus more on the processes and constraints but tend to overlook emotion.

## **Communicative Approaches**

Communicative theories see emotion as a communicative system which is used for communicating with other persons, one's mental state and as a way of asking for changes in others behavior [154].

### **2.1.4 Emotion Detection Approaches**

Emotion detection approaches are classified into four categories, lexicon-based, learning-based, a hybrid of both and other approaches not falling into the aforementioned three categories. Here, we look at these approaches and various researches conducted using these approaches

#### **2.1.4.1 Lexicon-based Approaches**

Here, emotion detection is approached as a keyword pattern matching problem. It is typically a problem of finding predefined keywords in a text document. These words have been grouped into various emotion categories such as happy, angry etc. The method typically consists of five steps where the input is a text document and the output is the emotion class. These five steps include tokenization, identification of emotion keywords, identification of intensity of emotion words, negation check, and then finally emotion classification [199, 235, 112]. Thelwall et al [265] in their work proposed SentiStrength, an unsupervised system for emotion and sentiment analysis which utilizes snippet-based and word-matching rules for detection, and handles intensifiers and negation. Various works have been done using this method with different algorithms. Lexicons like NRC [172] and Wordnet-Affect [248] are frequently used, either alone or in collaboration with other lexicons.

**Discussion:** This type of approach seems to have high accuracy rate as shown in various related publications but as was observed in various works that lexicon-based emotion detection methods have a few limitations [288, 235, 199] such as:

- ***Ambiguity in the definition of emotion keywords:*** Using emotion keywords is a direct way



to identify associated emotions. However, the meanings of words can be diverse, ambiguous and most often differ in meaning based on usage. In sentences such as ironic, sarcastic or cynical sentences [199, 112, 235], these words could have a different meaning and thus have a different emotion classifications.

- ***Failure of detecting emotion/sentiment of text without emotion-bearing words:*** Most of the techniques rely heavily on the emotion keywords. Thus, sentences which do not have any of the predefined keywords are assumed to be without emotion or neutral, and this is not always true [199, 112, 235]. For instance, “I passed my qualifying exam today” and “Hooray! I passed my qualifying exam today” should be categorised with the same emotion. However, the absence of “hooray” in the first example makes it impossible to detect the underlying emotion using lexicon-based approaches.
- ***Ignorance of linguistic information:*** The expression of emotion is influenced by syntactic structures and semantics. For example, “I laughed at him” and “He laughed at me” should have different emotions based on the speakers perspectives. Not considering linguistic information raises a problem in most lexicon-based systems [199, 112, 235].

An extension of keyword spotting approach is to assign a probabilistic affinity for specific emotion to words. These word probabilities can be derived from linguistic corpora and are often biased toward the genre of texts in that specific corpus. However, such an approach only considers emotional content at the word-level [199, 235, 112].

#### **2.1.4.2 Learning-based Approaches**

This type of approaches looks at the problem differently. It casts the emotion detection problem as one of text classification. That is, a classifier is trained to determine if a text document falls in any of the predefined emotion categories. This is different from the lexicon-based approach, as it tries to detect emotions using a classifier which is trained with labeled or unlabeled data to determine the emotion category of an input text [199, 292, 262, 235, 112, 292, 262]. More recent work on emotion detection using learning based approach can be further classified into unsupervised, semi-supervised and supervised learning approaches. These approaches often utilize one of the following machine learning techniques:

- **Naive Bayes Algorithm**(NB) : This is kind of classifier that uses Bayes Theorem [133]. Bayes Theorem works on conditional probability, which is the probability that an event will occur, given that another event has already occurred. Thus, the probability of an event can be calculated based on prior knowledge [133]. It assumes that all the features are unrelated/independent to each other. Naive Bayes is a fast, scalable and useful for binary and multi-class classification. Its biggest drawback is that it considers all the features to be unrelated/independent, this results in an inability to learn the important relationships between features [133].
- **Support Vector Machines** (SVM): In SVM, the training data is plotted in given dimensional space [110]. A binary SVM predicts a straight hyperplane dividing these 2 classes. The hyperplane is drawn by maximizing the distance between hyperplane and the nearest data point of each class [110]. This is called the maximum-margin hyperplane. The subset of the training data instances which are used to determine the hyperplane in SVM are known as the support vectors. The dimensionality of feature/attribute space does not have an effect on the performance of SVM classifiers [110].
- **Logistic Regression Method**: This uses a black box (Softmax function) function to understand the relation between the categorical dependent variable and the independent variables [164]. A dependent variable is the class to be predicted, while independent variables are the features/attributes used for predicting the target class. A Softmax function output values are always in the range of (0, 1) and their sum will always be equal to 1 [164].
- **Neural Networks**: Neural networks (NN) can be viewed as a complex function with numeric inputs and outputs [98]. The output values are determined by a combination of the input values, the hidden and output activation functions, and the weights/bias. A neural network with  $i$  inputs,  $h$  hidden nodes, and  $o$  outputs has  $(i * h) + h + (h * o) + o$  weights and biases. Neural networks are often trained using back propagation [98]. The aim of training a neural network is to find weights/ biases values, in other for the computed network outputs to closely match the training sets known outputs [98].

### 2.1.4.3 Supervised Learning

Supervised learning put simply, is basically learning from examples. Here, two sets of data are provided to the algorithm, training, and a testing set. The learner has to “learn” from the examples in the training set which are labeled, to enable it to identify unlabeled test set examples with high accuracy where possible [97]. Mohammad [166] conducted emotion detection experiments with Logistic Regression and Support Vector Machines (SVM). They used binary features which represented the presence/absence of unigrams and bigrams identified in the dataset. Alm et al [6] worked on fairy tales dataset and used Sparse Network of Windows (SNoW) learning architecture which is tailored for large-scale learning for the multi-class classifier. They experimented with different feature combinations and achieved an average accuracy of 0.63. Strapparava et al [251], carried out experiments focused on news headlines taken from news media websites and blog posts using Naive Bayes classifier and Latent Semantic Analysis (LSA). Gliozzo et al [84] used a type of Latent Semantic Analysis (LSA) which took into account the weighting schema of *tf-idf*. They carried out corpus-based experiments of posts from LiveJournal.com. The posts were mapped using mood to six emotion categories. Alm et al [6] used a manually annotated data set and explored automatic classification of children’s fairy tales sentences into basic emotions as defined by Ekman [68].

Gupta et al [87], performed emotion detections on customer service emails. A total of 620 email messages were used for training and 457 emails as test samples selected at random from a two million email corpus. They identified 8 reaction categories in the dataset such as business threats, report threats, legal threats etc. They experimented with Boosting [224] and Support Vector Machines (SVM). They used BoosTexter [225], a variation of AdaBoost which takes words and n-grams as features used for emotion classification. AdaBoost is a meta-learning algorithm which takes the outputs from many weak learners as a weighted sum to produce the final classification output. They archived an F-measure of 0.746, having used unigram features and orthographical features [87].

In Mohammad et al [167], worked with tweets. Each tweet had a hashtag corresponding to Ekman’s six emotions: disgust, anger, happy, fear, sadness, and surprise. They ignored tweets that have a hashtag in the middle, do not contain a hashtag and all retweets. The experiments carried out with supervised Support Vector Machine (SVM) algorithm was especially informative since

it showed that when the hashtags are compared to human judge results, they correspond to the annotations [167]. They found that the results were improved by combining labeled data in other domains with the Twitter corpus.

Vo et al [277], compared support vector machine(SVM) and Naive Bayes algorithms on a Twitter dataset containing tracking and updates on four earthquakes. They used hashtags to detect tweets that were about the earthquakes. They collected and annotated the tweet data for calm, unpleasantness, sadness, anxiety, fear, and relief emotions. Using supervised Multinomial Naive Bayes and Sequential Minimal Optimization (SMO), they created models for detecting the annotated emotions, with the best results produced by Naive Bayes [277].

In Li et al [137], their work was based on the theory that emotions triggering event are important. A crucial step in their approach is the extraction of emotion cause. They examined the dataset and constructed an automatic rule-based system to extract the emotion causes in the posts. They used a corpus of Chinese micro-blogs labeled by human annotators of 16485 posts classified into 6 emotions and neutral. Using 75% of the dataset as training set and 25% for the test set, they made use of a version of SVM proposed by Drucker et al (SVR) support vector regression [67], which uses SVM principles but returns a real number [67]. The classifier is trained based on extracted cause events to classify emotions in micro-blog posts.

Roberts et al [221] worked on experiments to detect emotions from suicide notes. These suicide notes were annotated by the authors and used for the experiments. They used for their baseline method for emotion detection an approach proposed by Robert and Harabagiu [220]. First, they use independent binary SVM classifiers, with each classifier working on individual emotion categories, they then create a single multi-label classifier by combining the individual SVM classifiers. This allowed the tweets to be classified into multiple emotion categories. They used the implementation of SVM found in WEKA [90, 221].

In Kunneman et al [124], their approach was training a classifier that will result in a balanced binary classifier with an equal amount of tweets with emotion hashtags and random counter-examples. They used Balanced Winnow algorithm [139] and found that 12 of the 24 hashtags tested had an AUC score of about 0.80 or higher and a precision score of 0.70 on two of the 24 hashtags [124].

#### **2.1.4.4 Semi-supervised Learning**

This type of learning uses the techniques of supervised learning for training but with a combination of labeled data and unlabeled data. This usually consists of a large quantity of unlabeled data and a small quantity of labeled data [97].

Tsur et al [269] worked on a product review dataset and used a semi-supervised approach using a strategy similar to the concept of k-nearest neighbors (kNN), to detect the presence of sarcasm in products reviews. They used features such as content words, frequent words or punctuation marks, to represent sarcastic texts, with results of high precision and recall scores.

Haggag used a semantic frame approach for detecting emotions in text [89]. By using a knowledge base, they stored semantic and syntactic information. Learning is provided by training a Knowledge-Based Artificial Neural Network (KBANN) to aid classification decision of detecting emotions in text. Emotion detection is done by matching against the entries in the knowledge base. The dataset had the following emotion categories: sad, happy, anger, joy, disgust, fear, astonishment and natural, also comprised of short paragraphs of about 3 sentences with 600 labeled with positive emotion and 400 negative emotion. Their model had higher accuracy when compared to the linguistic, pragmatic and keyword models.

In Bann et al [30], they collected Twitter data of tweets that contain emotion keywords related to Ekman's six basic emotions [68]. They used Document-Emotion Latent Semantic Algorithmic Reducer (DELSAR) which uses emotion set and clusters documents emotion clusters. They also make use of the Emotion Profile concept and conduct analysis of compound emotions [30].

#### **2.1.4.5 Unsupervised Learning**

In unsupervised learning, a type of machine learning that draw inferences from a datasets input data in the absence of labeled responses, basically learning without a teacher.

Calvo et al [46] introduced a way of using normative databases for processing text and compare it with categorical approaches. For their experiments, they utilized a lexicon and a bag-of-words approach to build/create sentence vectors. They then applied Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA) and Non-negative Matrix Factorization (NMF) to their dataset. Their experiments show that the categorical model with NMF performed best along with the dimensional model [46].

Tang et al [257] combined emotion lexicon, emoticons and emotion-shifting rules to detect emotion in tweets. They use an unsupervised method to balance the effectiveness and the scalability. They manually annotated data on earthquakes and conducted experiments whose results show that the precision reaches 80% [257].

Agrawal et al [5] presented a context-based unsupervised approach for sentence level emotion detection. Their approach does not depend directly on emotion lexicons such as WordNet-Affect, rather refer to emotion-bearing words as NAVA (Noun, Adjective, Verb, Adverb) words. First, they extract words considered as emotion bearing words. Using part-of-speech tagging, they extract all words in an input sentence that fall into the NAVA words tagging by first assuming that all emotion-bearing are NAVA words. They use a set of affiliated emotion words in order to calculate the emotion vector of a given emotion bearing word. This is done by calculating the semantic relatedness score of the word and the emotion concept score generated from the Point-wise Mutual Information (PMI) score of the co-occurrence of the NAVA word and the affiliate emotion word [5].

#### 2.1.4.6 Discussion

The algorithms in this category have very good performance and accuracy levels as shown in various related publications. However, it has a few pitfalls such as:

- **Identifying Emotion Indicators:** Learning-based methods tend to inherit the problems of lexicon approach as they generate probabilities based on words in the input text. The classifiers require features such as words and emoticons to assist in the detection of an authors emotion [199, 112, 235].
- **Dependence of labelled data for training:** This type of approach require large training data, and this data has to be manually labeled. It is both time consuming and labour intensive to label data manually, and this labeling has to be done for each application domain. Domain specific data results in classifiers performing well in their trained domain and poorly on different domains [112, 113, 257, 199].
- Most of the algorithms studied also show a lack of use of linguistic information to aid in emotion identification, problems in detecting emotion cues, and sometimes biased toward the corpus-specific genre of texts [199, 112].

- The algorithms in this approach tend to stick to Ekman's 6 basic emotions [68] resulting in over-simplified emotion categories and emotions outside these categories classified as neutral (which they are not) or ignored completely.

#### **2.1.4.7 Hybrid Approaches**

Given the flaws of keyword-based and learning-based methods, some researchers adopt a hybrid approach. This is most often the combination of keyword-based and learning-based methods. This combination helps to improve the level of accuracy generated. In Wu et al [288], they use a rule-based method to extract semantics of various emotions in text as well as lexicon ontology features. These features are used in the training of their learning module. This method though with limited emotion categories outperforms previous approaches [199, 235, 112].

Aman et al [10] used an annotated dataset of blog post sentence corpus, which contained information about emotion category, intensity, and emotion indicators. They found that instances of fear and happiness had high annotator agreement, as well as sentences with high emotional intensity. They used General Inquirer [247], WordNet-Affect [248], and a combination of both with Support Vector Machines (SVM) and Naive Bayes. The best performance was a significantly high accuracy of 73.89%.

**Discussion** Hybrid approaches perform with a reasonable level of accuracy. However, it appears to inherit both the pros and cons of learning and keyword-based approaches.

#### **2.1.4.8 Other Detection Approaches**

Neviarouskaya et al [182] used a rule-based method for detecting Ekman's primary emotions [40] in data set of blog post sentences. Liu et al in their work used real-world knowledge from a common-sense knowledge base for sentence level emotion detection. First, they extract from the knowledge base sentences with emotion, which are then used for building emotion models for labeling sentences with the corresponding emotions [68].

Singh et al [237] described sentence level emotion detection using supervised learning approach and tagged dictionary and unsupervised learning approach using ISEAR database. They use a PCFG parser, which is a probabilistic version of a CFG (context-free grammar) where each production has a probability and is an important part in training machine. The algorithm used is Probabilistic Latent

Semantic Analysis (pLSA), which is a technique mostly used in topic modeling. It discovers underlying semantic structure in data by modeling the co-occurrence information under a probabilistic framework. The algorithm views data in terms of documents, words, and topics these are its main variables.

Das et al [64] used preprocessed WordNet Affect lists for semi-automatic word level emotion annotation process. Using Conditional Random Field (CRF) for word level based emotion detection, they produced an overall accuracy score of 87.65% with a dataset of 250 sentences. Gill et al [81] used a Latent Semantic Analysis (LSA) and Linguistic Inquiry and Word Count (LIWC) method for determining basic emotions [68] in data set of sentences taken from blog posts [81].

### **2.1.5 Implicit Emotion Detection**

The expression of emotion in written text is through the use of words and most often emotion-bearing words such as “happy”. However, emotions can be adequately expressed without the use of emotion-bearing words. For example, given two sentences “The outcome of my exam makes me happy.” and “I passed my exam.”, both sentences express the emotion of happiness, with the first expressing it explicitly and the second implying it. Most research in the area of emotion detection focuses on explicit emotion detection [235, 154]. Implicit emotion detection is a much more difficult task and the approaches which rely on emotion lexicons are inapplicable here. Although it is possible to train supervised classifiers from annotated data, acquiring sufficient annotated data for training requires heavy manual effort. Implicit emotion detection in text refers to identifying and classifying text into emotion categories in the absence of emotion-bearing words.

Liu et al [142] argued that real-world knowledge of everyday emotional situations (e.g., “getting into a car accident”) can be used to detect implicit emotions. They made use of the Open Mind Common Sense Corpus (OMCS) [238] which holds 400,000 facts about the everyday world. From OMCS, they extracted sentences which contain affect based on Ortony’s Affect Lexicon [191]. Those affect keywords from the lexicon were pre-classified into Ekman’s basic emotions and acted as “emotion grounds” which propagated their value to concepts related to commonsense relations. Based on the syntactic parse results obtained from text, they created four linguistic models coupled with some hand-crafted rules to classify sentences into one or more of Ekman’s basic emotion categories. Their approach is restricted by the knowledge source and the affect lexicon used. Also,



the approach largely depends on the subject-verb-object-object structure of sentences which is not always present.

Balahur et al [24] constructed the EmotiNet ontology which started with three knowledge cores including kinship relations, emotions, and actions. The EmotiNet ontology was then extended and populated using the situations found in International Survey On Emotion Antecedents And Reactions (ISEAR) dataset<sup>1</sup>, and was further expanded using existing resources and other commonsense knowledge bases. EmotiNet holds real-world situations and emotional reactions in which commonsense is key to the interpretation of affect. In subsequent work, EmotiNet was extended to enable implicit emotion detection [25]. The problem with the EmotiNet approach is that it is domain-dependent, since the approach would fail if the knowledge on certain actions is not defined in EmotiNet.

There are also some other approaches not relying on commonsense knowledge, but can still be used for implicit emotion detection. For example, Agrawal and An [5] already discussed above, although this approach did not directly target at implicit emotion detection, it occasionally identified emotions of sentences without emotion-bearing words. Nevertheless, their approach depends on the external knowledge sources such as Wikipedia for the calculation of PMI values which is computationally expensive.

**Discussion:** Most of the approaches we encountered for implicit emotion detection in text rely on some external knowledge sources. However, the use of knowledge sources is not robust enough to be applicable in any real-world context as it is impossible to cover all possible entity specifications. Another limitation is that it does not give consideration to semantic relationships of linguistic components in the given sentence and the contextual meanings of words in the sentence.

### 2.1.6 Summary on Emotion Detection

Given the progress and challenges in this field of study as presented above, the research work reported in this thesis aims to address the following limitations of existing approaches to emotion detection from text.

- **The lack of labeled data:** We propose an approach for implicit emotion detection which will be able to classify text into emotion categories without the need for labeled data.

---

<sup>1</sup><http://www.unige.ch/fapse/emotion/databanks/isear.html>

- **Dependence on emotion lexicons:** We propose an approach for implicit emotion detection, which will not have a dependency on emotion lexicons.
- **Feature engineering perspective:** Our proposed approach presents a set of rules used to allocate values to the model variables and reduces the amount of feature engineering required.
- **Over-simplified emotion categories:** Our proposed solution utilizes the OCC model which has 22 emotions, with emotions such as sorry-for and happy-for which indicate the direction of the emotion, rather than classifying all emotions under the same category regardless of the direction (towards self or towards others) of said emotion.

Our approach for the detection of blame and praise is discussed in detail in chapter 3 of this thesis.

## 2.2 Blame and Praise Detection

With the prevalence of social networking sites, the continual growth of social computing, there is a need for more and more systems which utilise human insight for computer-related tasks [196]. In recent years, social computing is becoming key in many research areas and technological systems such as learning, human-computer interaction, entertainment and many more. A major aspect of social computing is the ability of an entity to infer the social behaviour of not just itself but of other entities as well [196, 151]. This inference includes the ability to pass judgment and determine if an entity is blameworthy or praiseworthy and allocate blame or praise where appropriate [267]. Blame and praise are closely related, to blame an entity is to hold that entity morally responsible for doing something of a negative outcome while praise is to hold that entity morally responsible for doing something of a positive outcome [69]. The volume of text data in recent years is continuously growing, so is the need to understand the content of the data. 90 percent of today's online communications is in text format [58]. Detection of blames can be used in a variety of applications such as identifying entities holding moral responsibilities in multi-agent systems, in emotion detection from text in helping with the identification of emotions such as guilt, remorse, admiration, shame [191], in forensic linguistics for legal cases, and in automatic identification of relief efforts by crawling the web during a disaster and many more.

### 2.2.1 Theories of Blame

Blame is in the family of “moral judgments”. It deals with evaluating agents for their involvement in events to determine if the agent is blameworthy or praiseworthy. As discussed in [145], in the family of moral judgments one needs to distinguish at least three types :

1. Setting and attesting to norms for example, avowing one norm as overriding another or stating an imperative.
2. Evaluating events, the outcome of events and behaviors in relation to norms, e.g. judging an event as bad or good.
3. Performing agent evaluations to see their involvement in norm-related events, for example judging someone as blameworthy or morally responsible.

The key difference between the three types of judgment includes the following: *Type 1* is directly involved with norms, while *Types 2* and *3* are judgments through evaluation in relation to those norms. Furthermore, *Type 2* focuses on events, while *Type 3* focuses on agents. Blame falls into the category of *Type 2* and *3* [145]. To find an entity as blameworthy, one is holding that said entity morally responsible for doing something wrong. Blame is cognitive in relation to the process that leads to the judgment of blame; blame is also social in relation to the act of showing a judgment of blame to a different entity.

There is a variety of theories of blame, these theories could be organised in different dimensions depending on the purpose. We consider here just two dimensions:

- First, we could categorise blame theories based on its blaming content [145]. This dimension covers theories that believe that blame is found in judgments of ill will and the theories that state that blame can be said to be a response to ill will which is emotional [145, 267].
- We could categorise blame theories based on psychological states which are in line with blame. These would include: *Cognitive* theories where blame is viewed as an evaluation or judgement an entity makes about an agent in relation to attitudes or actions; *Conative* theories where blame is viewed with focus on aspect considered essential to blame the motivational aspects such as intention; *Emotional* theories see blame as an emotional expression; and

*Functional* accounts identify blame by its functional role and can be more flexible than other three categories [267].

### **2.2.1.1 Cognitive Theories of Blame**

These theories view blame as a judgment one makes about an agent based on their action. According to Smart [240], it is a form of grading which is no different from how one decides if an orange is better than another orange. He states that there is more to blame than mere evaluation of an agent's actions, blame implies responsibility of a negative evaluative judgment. The work of Gary Watson [283] suggests that blaming someone is to say that the person is responsible for their actions and have failed to meet a given criteria. In these theories, there is no need for the one passing judgement to be conatively or emotionally exercised. These two theories are closely related but have differences such as Smart theory believes blame requires moral responsibility, while Watson's theory sees moral responsibility as a virtue of blame and one of many by which we can blame others for their actions. Pamela Hieronymi [99] articulates cognitive accounts of blame, an evaluation of the ill will shown by a person. Our view or rating of other people's judgments about our will ill or not, can carry a level of blame. A potential problem of cognitive theories of blame is that there is risk conflating blaming with judging blameworthy. According to Hieronymi [99], the fact that said judgment could elicit such different responses means that the judgment alone cannot constitute blame.

### **2.2.1.2 Emotional Theories of Blame**

Our attitudes and emotions are closely linked with our ability to function as morally responsible agents, which is needed for day to day relationships with others. This theory views others as morally responsible agents is an emotional response and not one of judgement [252].

According to Wallace [278] we hold people morally responsible as a form of response to the person's actions or as a way to find the appropriate response to the said action. For Wallace, blame is essentially implicated in the stance we take holding an individual responsible. It states that one can take a stance on "*holding responsible*" another person and not be emotionally exercised, but in order to actually blame an individual, one must be emotionally exercised by responding with emotions such as anger/resentment [278, 279].

George Sher [234] argues that one can blame without responding with negative emotional reactions, thus an emotional response in this view is not needed for blame. A good example of this is

when we blame a loved one.

### **2.2.1.3 Functional Theories of Blame**

Instead of identifying blame with judgment or emotion or combination of attitudes, functional accounts of blame state that blame plays a functional role, which comes across as a form of protest. Thus, when blaming others, we are upset about their actions/character. Hieronymi [99] and Talbert [256] argue that reactive attitudes such as resentment serve as powerful forms of protest. When one changes their attitude or stance, that change must serve a function which in the case of blame is a protest [243]. Michael McKenna [158, 159] argues that blame serves functions like disapproval and a form of communication.

There are some problems with this theory such as no clear indication that protest is independent of blame, in other to protest without appealing to blaming attitudes. The nature of blame is not clarified by using the notion of protest [267]. Second, protest seems to be more paradigmatically expressed and difficult to understand unexpressed protest [267].

### **2.2.1.4 Conative Theories of Blame**

Conative theories of blame focus on the elements like desires and intentions, this theory considers both as essential to blame. George Sher [234] is of the view that "when the cognitive component of judging blameworthy is accompanied by this desire, which reflects our general commitment to morality, then we are said to be blaming". Smith [242], opposes Sher's view on blame about the role played by desire. He presents a case of blame which is placed on a politician who led his country to war, even though the people do not desire this, it is unclear the role said desire to play on the act of blaming the said politician.

## **2.2.2 Blameworthiness**

One's action is blameworthy when they are found to be morally responsible for some wrongdoing. In contrast, they are praiseworthy for doing something right [145, 267]. The path model to blame helps to answer the question: "*When is it appropriate for X to blame Y?*". The answer is that "*Only when Y deserves it*". Hence in order for an agent to be blameworthy, certain conditions must be satisfied:

- **Moral Agency.** As we stated earlier, being to blame is not adequate for being blamewor-

thy. According to Gary Watson, elements like earthquakes are most often blamed for various things that happen, but earthquakes are not actually blameworthy because it cannot react effectively and competently in moral matters [69]. There is a wide acceptance that blameworthy agents must have the capacity to reason about and execute a decision, thus the agent must be a moral agent [267]. This means that entities such as earthquakes and floods cannot be moral agents.

- **Freedom.** When one thinks an individual is considered blameworthy if said individual exercises free will [69]. Often times comments like “*I couldn’t help it*” and “*I was forced to do it*” are excuses and are often enough to render blame inappropriate [267]. Free will is seen as a person’s capability to control, by process of selection chose one of two futures presented before one. Our vulnerability to coercion, situational pressures, and manipulation which robs us of our freedom to chose, most often provides us with an exemption from blame [69, 267].

### 2.2.3 Path Model of Blame

The “Path Model of Blame” proposed by Malle et al [145], the model expresses that inside of the theoretical structure currently in place, standard social cognition gives rise to blame judgments. Blame judgments involve information which is important to other concepts and verifying the meeting of various required criteria [69, 267]. Blame seems to be centered around events and outcomes [145, 267]. According to Malle et al [145], the model applies equally well to both events which are time-extended processes and outcomes which is the result of events.



Figure 2.6: The Path Model of Blame [145].

Figure 2.6 shows the “Path Model of Blame”. On the hierarchy of en route to blame, the logic has to proceed along particular paths, as represented in Figure 2.6. From the structure, blame emerges if the perceivers first detect an event has violated the perceived norm (*Event Detection*); after the detection of the negative event, it has to be identified that an agent caused the event (*Agent Causality*) [145]. If no (moral) agent has caused the negative event, blame cannot be established. The Path Model states that the causal involvement of an agent falls into two categories, either intentional or unintentional. On the intentional path, if the negative event in question is evaluated and found to be intentional, the perceiver must now consider reasons for this action (*Reasons*) [145]. Blame is present, but the degree of blame is dependent on the reasons. When an agent is found to have unintentionally caused the said event, the event perceiver considers the degree of *obligation* and *capacity* (including the capacity to foresee or foreknowledge of the event) of the agent to prevent the negative event. According to the Path Model, it is only when an agent is found to have both the obligation and capacity that the agent will be blamed for the negative actions [145]. Adding the capacity for practical reasoning to the power of free will, one ends up with a morally responsible agent. As social intelligence is becoming central in a number of information and communication systems, so is research in the area. Before starting this work, we found a lot

of work in agent causation detection in text but very few in the area of detection of blame/praise [145]. In this section, we discuss some articles which are related to our work. We did not find much research in the area of blame/praise detection in text. We however present here work which is close to what we are doing. Attribution Research looks at how one makes sense of the world by attributing behavior and events to their causes. It is basically ascribing a cause to an event as well as the judgements made [126]. There are other works not directly related to text processing but useful in understanding computational models for blame/praise detection in text. In Mao et al [151], they adopted the shaver model terminology and represented causal knowledge in hierarchies which allow a concise description of the causal relationship between events and states. An example was presented, using the online text data crawled from 25,103 web pages from news outlets related to Al-Qaeda [151]. A set of manually defined linguistic patterns and rules were used to extract actions and the action preconditions and effects which were then used to represent causal knowledge [151].

Although the research presented in [151] shed light on analyzing the causal relationship between events and states, their reliance on manually defined linguistic patterns for identification and extraction of actions and action precondition and effects limits the scope of study since it involved heavy manual effort. Their approach is also domain dependent and cannot be generalized to other application areas. In this chapter, we propose an approach for blame/praise detection from text.

#### **2.2.4 Summary on Blame and Praise Detection**

Given emotions such as guilt and admiration which often require the identification of blameworthiness and praiseworthiness as stated in the OCC model, we propose an approach for the detection of blame and praise in text. This approach is adopted from the “Path Model of Blame” focused for the identification of expressions of blame and praise in text. The correct identification of blame and praise in text helps us to achieve our greater goal of detecting expressions of implicit emotions in text. Our approach for the detection of blame and praise is discussed in detail in chapter 4 of this thesis.

### **2.3 Conflict of Interest Detection**

A key feature of Wiki sites is to allow people from all over the world to add or modify articles anonymously and without consequence [96]. This enables people with malicious intentions to use articles for promotions or to try and discredit certain products/services, organisations, or individuals



[2, 214].

Conflict of Interest (CoI) is defined as a situation in which a person or organisation is involved in multiple interests, financial interest, or otherwise; one of which could possibly corrupt the motivation of the individual or organisation<sup>2</sup>. According to Wikipedia, content on Wikipedia and other Wiki-media projects “must be written from a neutral point of view (NPOV), which means representing fairly, proportionately, and without bias. NPOV is a fundamental principle of Wikipedia and a policy which is non-negotiable”<sup>3</sup>. NPOV refers to representing all views in relation to this topic which are published in reliable sources in a neutral and unbiased way. It is especially important for articles on controversial issues, where a great variety of viewpoints and criticisms can be found. A neutral representation according to Wikipedia will differ points of views presented but not as widely accepted facts. To comply with the NPOV policy, editors must follow the principles below: “*avoid stating opinions as facts, avoid stating seriously contested assertions as facts, avoid stating facts as opinions, prefer nonjudgmental language, indicate the relative prominence of opposing views*”<sup>4</sup>. CoI editing happens when an editor contributes to Wikipedia about themselves or their relationships such as family, friends, clients, employers, and financial links, etc. Often times, CoI editing does not comply with NPOV. CoI editing is discouraged strongly by Wikipedia as it tends to undermines the confidence of the public in it, and causes public embarrassment to the individuals being promoted. It is easy to assume that CoI is just bias; however while it is not possible for CoI to exist without bias, bias can often exist in the absence of a CoI. One’s beliefs and desires can lead to biased editing, but that does not constitute a CoI.

The growth of Wikipedia makes it increasingly difficult for both Wikipedia users and administrators to manually monitor articles. Wikiscanner<sup>5</sup> or its open-source clone WikiWatchdog<sup>6</sup> were developed to identify edits that organisations made on Wikipedia by matching their IP addresses with anonymous article edits. This shows that various organisations or their staff edited or removed parts of Wikipedia articles that concern the organisation. In 2014, Sony Pictures Entertainment confidential data were released online which contain thousands of emails and documents. Some documents reveal that Sony employees edited Wikipedia pages relating to Sony Pictures Entertain-

---

<sup>2</sup>[https://en.wikipedia.org/wiki/Conflict\\_of\\_interest](https://en.wikipedia.org/wiki/Conflict_of_interest)

<sup>3</sup>[https://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

<sup>4</sup>[https://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

<sup>5</sup><https://en.wikipedia.org/wiki/WikiScanner>

<sup>6</sup><http://wikiwatchdog.com/>

ment and also Sony paid editors to edit their pages<sup>7</sup>. This shows that various organisations may be doing this either to boost themselves or to play down their competitors. Bhosale et al. [34] presented work on detecting promotional content in Wikipedia. They looked at the content features, structural features, network features, edit history features, overall sentiment score, trigram language models and PCFG language models. They found that the stylometric features influenced results the most. Chandy et al [50] proposed Wikiganda which used article controversy indicators such as number of revision and number of unique editors and revisions-level metrics such as vandalism, sentiment, conflict of interest detected based on IP address of editors, etc., to identify propaganda in Wikipedia articles.

When an edit is made on Wikipedia, the editor can either register for an account or edit anonymously. When done anonymously, Wikipedia uses the IP address to identify and distinguish the article instead of a username. WikiScanner or WikiWatchdog listed “anonymous” edits related to real-world organisations. They work by comparing a list of all IP addresses that have made edits to Wikipedia with IP addresses which belong to real-world organisations and returning a list of “anonymously” edited articles made from the organisations’ IP addresses. Although WikiScanner or WikiWatchdog can be potentially used for CoI detection, they suffer from a number of limitations, for example, they don’t analyse the content itself and don’t consider edits done by registered users. Also, to avoid the detection by WikiScanner or WikiWatchdog, one would simply make an edit from a IP address not belonging to a real-world organisation.

Apart from work relating to vandalism, bias or CoI detection, there has also been some research focusing on developing a systemic-constructivist approach for knowledge construction in Wikipedia. In particular, it was found that the NPOV policy and other Wikipedia rules are important in supporting Wikipedia editors in removing biased phrasing, personal speculations and opinions [188].

The problem of content-based CoI has never been investigated before. We believe that our work will inspire further development of automated systems for CoI detection based on text content. However, CoI is closely related to other NPOV-disputes such as Bias and Vandalism. We take a closer look at research work in these closely related areas.

---

<sup>7</sup>[https://en.wikipedia.org/wiki/Wikipedia:Wikipedia\\_Signpost/2015-04-22/Special\\_report](https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-04-22/Special_report)

### 2.3.1 Bias

The growth of the internet means that online reference sites play a big role in shaping the view of the public and understanding the inner workings of these systems have become important for society and businesses. This area has gained ground in recent years with researchers looking at subjective components such as detection of sentiments, emphasis or perspective and topic categorization, all of which has reached a reliable level of performance.

News website reports are generally expected to have an impartial depiction of facts. However, there are some cases news website present news based on their own standpoints, sentiments or interests. These biased reports are one-sided information that most often lead to wrong judgments and opinions. Since such outlets want to appear objective, journalists avoid using opinionated vocabulary, but sometimes use other means such as cherry-picking facts or quoting how other persons feel as a means to express their opinion. Thus identifying sentiment expressed in such a way can be difficult. Most of the work done in the area of bias detection is centered on bias in news articles. Balahur et al [29] realized the difference in news opinion mining and other text types, and described subtasks for news opinion mining: what is the target; separating news content from its polar sentiment directed at the target; and clear explicit opinion. They also identified and made a distinction between possible views of a news article author's view, reader's view and the text. Their work addressed these three views. During analysis, the author may communicate opinion, by either omitting/stressing parts of the text, from the reader's viewpoint, the text can be highly subjective based on interpretations and can be swayed by culture, religion and other factors. The experiment results showed that focusing on the text content produces a better performance.

Zhang et al [297] focused on the sentiment in news articles and developed a system which can detect and visualize sentiment of websites. The system can extract subtopics based on a topic and subtopics level sentiment differences. The sentiment difference in topics of websites can aid users in news credibility decisions. Experimental evaluations of their approach showed good accuracy of sentiment extraction and subtopic extraction. Recasens et al [214] in their work, analyzed human edits meant to correct biased articles. They identified that there are two bias classes: framing bias, like praising which can be linked to subjectivity and epistemological bias. They find that the system performs rather well and that features based on subjectivity and sentiment lexicons are very

helpful in detecting bias. Other research work in this area builds upon previous works in areas such as subjectivity detection and stance recognition in texts, varies in choices of semantico-syntactic elements. Given the undesirable nature of bias in reference works, the expression of bias in this area tend to be more implicit and thus harder to identify by both humans and computers algorithms.

### 2.3.2 Vandalism

Wikipedia being an open community can allow anyone to participate and contribute, this ability became the essence and central to its success. This openness also creates problems which tend to endanger overall growth of the Wikipedia project [210, 96]. One major problem is that some users target Wikipedia for malicious purposes, which impacts the encyclopedia negatively [281, 96]. Vandalism can be defined as any modification of content with a cautious effort to put the integrity of Wikipedia into question [281, 96, 210]. This definition makes the concept of vandalism highly subjective. With the continual growth of Wikipedia, it becomes much harder for Wikipedia users and administrators to police articles manually [53]. Given cost of vandalism to Wikipedia, it is not surprising that various attempts for vandalism detection already exist and has given rise to various research activities to understand and prevent vandalism. Early tools for vandalism detection used handcrafted rules with encoding heuristic vandalism patterns for labelling vandalism articles, such bots include ClueBot<sup>8</sup>, MartinBot<sup>9</sup> etc. These Bots typical rules were narrow and included items like: the amount of text added or removed, capital letter ratio, and the presence of vulgarisms. Potthast et al [210] manually crafted a feature set using various content-level properties and meta information. They trained a logistic regression classifier with their dataset. With 0.83 precision and 0.77 recall and when compared to current systems used in Wikipedia, their approach was faster and outperformed in F-Measure by 49%.

Chin et al [53] constructed a statistical language models for articles from their revision history. This is an alternate view of content edit and stems from the understanding that non vandal content belongs together. According to their approach, if bad content is added to edit, its compression level is found to be lower than it would be if the text is related to the existing content. The approach presents a flaw that huge edits are often tagged as vandalism regardless of content quality. West et al [284] used reputations on users and edits. They showed that the use of meta-data is helpful

---

<sup>8</sup><https://en.Wikipedia.org/wiki/User:ClueBot>

<sup>9</sup><https://en.Wikipedia.org/wiki/User:MartinBot>

for vandalism detection, and that there are more possible indicators of vandalism that cannot be handcrafted. Wang and McKeown [281] present a linguistic approach based on shallow syntactic patterns. Potthast et al [211] presented a survey of the various features that have been employed for this task of vandalism detection.

Manoj et al [96] explored linguistically motivated approaches for detection of vandalism and suggested that vandalism is a unique genre which constitutes of people with similar linguistic behavior. Their experimental results show that statistical models provide proof of this unique vandalism language styles.

### **2.3.3 Summary on Conflict of Interest Detection**

In this section, we looked at Conflict of Interest (CoI) detection in Wikipedia. We identified that CoI falls within the NPOV-dispute category of Wikipedia. Because there are no major work in content-based CoI detection we have reviewed other works within its parent category such as Bias and Vandalism to aid in identifying possible features and approaches that can aid with CoI detection in articles. We have identified various forms of expressing bias and identified stylometric features used in expressing vandalism, these will feed into our feature engineering process for identifying best features for Conflict of Interest (CoI) detection in Wikipedia in subsequent sections.

## **2.4 Summary**

In this chapter, we discussed emotion detection, its theories, models and computational approaches. We introduced models and discussed related work in the emotion detection, blame/praise detection and conflict of interest detection. In subsequent chapters we will present our approaches with the models introduced in this chapter and present our experimental results and evaluations.

## Chapter 3

# A Rule-Based Approach to Implicit Emotion Detection in Text

In written text, emotions can be expressed explicitly or implicitly. Various research in written text based emotion detection is often focused on explicit expressions of emotions in text. In this chapter, we present a rule-based pipeline approach for detecting implicit emotions in written text( detecting emotion in the absence of emotion-bearing words) based on the OCC Model. We have evaluated our approach on three different datasets against five emotion categories.

### 3.1 Introduction

Researchers agree on some underlying factors of emotion which include that emotions are elicited in reactions to our environment, they are subjective, they involve an appraisal process and they involve physiological reactions. Our motivation for using the OCC model is because it defines emotions as valenced reactions to events, agents and objects and uses valenced reactions as a means to distinguish between emotions and non-emotions. It also classifies emotion based on experiencer. As such, the OCC model provides an opportunity for applying the NLP techniques for the identification of emotion-inducing situations, the cognitive state of users (expressed usually as adverbs and adjectives), and the variables causing emotions [228]. Particularly, the OCC model identifies different emotions depending on the direction of the emotion. This overcomes the over-simplification of emotion categories problem most approaches have.

For example, for the sentence “I passed my exam.”, the OCC model will return emotions of self such as *Joy*. For another sentence “He passed his exam.”, the model will return emotions of others such as *Happy-For* as opposed to simply returning *Joy* regardless of the direction

of the emotion.

We also considered working with the PAD model but our motivation not to use the PAD model is primarily because the PAD model has varying definitions and interpretations of Pleasure, Arousal and Dominance [18]. We believe the PAD model was more suited for systems that required a continuous emotional state of the subject than for text-based emotion detection which is just a snapshot of an individual's state at given point in time [18]. There is also a lack of testing and training data available for research and set up for baseline systems for the PAD model.

## 3.2 Our Approach

In order to use the OCC model for emotion detection, we need to first assign values to a list of variables defined in OCC, and then use a set of pre-defined rules to identify an emotion for a given text. We focus on identifying emotion in relation to events and actions only and leave the detection of emotions associated with objects as future work. The list of rules is shown in Table 3.1. For example, the first row of Table 3.1(a) can be read as

**If** Direction = "Self" **and** Tense = "Future" **and** Overall Polarity = "Positive" **and** Event Polarity = "Positive", **then** Emotion = "Hope".

In this section, we describe how we assign values to various OCC variables. Here, the OCC variables correspond to the set of specific rules that can be used to identify different emotional responses. We are specifically interested in detecting implicit emotions from text where there are no emotion-bearing words. It is worth noting that emotion-bearing words are different from polarity-bearing words. An emotion-bearing word can be described as words which on their own can convey emotions. For example, the word "passionate" can convey an emotion of *Joy*. Polarity-bearing words, on the contrary, express *positive* or *negative* polarity in a given context. For example, the word "pass" expresses a *positive* polarity as in "I passed my exam.". But the word "pass" does not have an explicit prior emotion associated with it. Hence, it is more likely that emotions-bearing words also have a polarity, but not all polarity words convey specific emotions.

### 3.2.1 Overall Steps

We first perform pre-processing on text in order to be able to assign values to the OCC variables. For pre-processing we carried out, we list the pre-processing steps below.

Table 3.1: Rules for emotion detection.

<b>Direction</b>	<b>Tense</b>	<b>Input Variables</b>		<b>Output Emotion</b>
		<b>Overall polarity</b>	<b>Event polarity</b>	
Self	Future	Positive	Positive	Hope
Self	Future	Negative	Negative	Fear
Self	Present	Positive	Positive	Joy
Self	Present	Negative	Negative	Distress
Self	Past	Positive	Positive	Satisfaction
Self	Past	Negative	Negative	Fears-confirmed
Self	Past	Positive	Negative	Relief
Self	Past	Negative	Positive	Disappointment
Other	All	Positive	Positive	Happy-for
Other	All	Negative	Positive	Resentment
Other	All	Positive	Negative	Gloating
Other	All	Negative	Negative	Sorry-for

(a) Event-based.

<b>Input Variables</b>		<b>Output</b>
<b>Direction</b>	<b>Polarity</b>	<b>Emotion</b>
Self	Positive	Pride
Self	Negative	Shame
Other	Positive	Admiration
Other	Negative	Reproach

(b) Action-based.

<b>Input Variables</b>		<b>Output</b>
<b>Event</b>	<b>Action</b>	<b>Emotion</b>
Joy	Pride	Gratification
Distress	Shame	Remorse
Joy	Admiration	Gratitude
Distress	Reproach	Anger

(c) Compound emotions.

- *Sentence splitting and tokenisation.* We use the Stanford Tokenizer<sup>1</sup> to split the text into sentences and then further split each sentence into a sequence of tokens, which roughly correspond to “words”.
- *Part-of-Speech Tagging.* We assign a part-of-speech tag such as nouns and verbs, etc., to each individual word in a sentence using the Stanford part-of-speech tagger<sup>2</sup>.
- *Word Sense Disambiguation (WSD).* Words with more than one sense could have different meanings in different contexts. We used NLTK (Natural Language Toolkit)<sup>3</sup> which performs the classic Lesk algorithm for WSD. The Lesk algorithm for WSD, works by looking at the sense definitions of each sense for the individual words in a sentence. Each word in the sentence is assigned the sense which has the maximum number of definition words in common with the sense definitions of other words in the sentence.

<sup>1</sup><http://nlp.stanford.edu/software/tokenizer.shtml>

<sup>2</sup><http://nlp.stanford.edu/software/tagger.shtml>

<sup>3</sup><http://www.nltk.org/>



- Assume we have a sentence

$$W_1W_2W_3$$

(where W is a word)

- Each word in our sentence has senses or synsets

$$W_1 : (W_{1s1}, W_{1s2}, W_{1s3}), W_2 : (W_{2s1}), W_3 : (W_{3s1}, W_{3s2})$$

- Creating a combination of the senses of all the words in the sentence:

$$(W_{1s1}, W_{2s1}, W_{3s1}), (W_{1s1}, W_{2s1}, W_{3s2}), \dots (W_{1s3}, W_{2s1}, W_{3s2})$$

- Iterate through the sense combination and return the combination that provides maximum score of overlap of words.
- *Dependency Parsing.* This process is to provide a representation of grammatical relations among words which are present in a sentence. The Stanford dependency parser<sup>4</sup> is used to extract textual relations.
- *Subject, Object, Event and Action Detection.* We need to identify events and actions from sentences. It is possible that multiple events can be expressed in a single sentence and similarly for actions and emotions. The subject(s) and object(s) of a sentence can be identified from the dependency parse results. Based on the POS tagging results, we identify noun phrases as events and verb phrases as actions. We also perform a passive test to determine if a sentence is in an active voice or passive voice. This will be useful in identifying the direction of emotion.

---

<sup>4</sup><http://nlp.stanford.edu/software/stanford-dependencies.shtml>

## Stanford Parser Output

*I passed an exam that I was absolutely certain that I had failed.*

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBD passed)
      (NP (DT an) (NN exam))
      (SBAR (IN that)
        (S
          (NP (PRP I))
          (VP (VBD was)
            (ADJP (RB absolutely) (JJ certain))
            (SBAR (IN that)
              (S
                (NP (PRP I))
                (VP (VBD had)
                  (VP (VBN failed))))))))))
    nsubj(passed-2, I-1)
    root(ROOT-0, passed-2)
    det(exam-4, an-3)
    dobj(passed-2, exam-4)
    mark(certain-9, that-5)
    nsubj(certain-9, I-6)
    cop(certain-9, was-7)
    advmod(certain-9, absolutely-8)
    ccomp(passed-2, certain-9)
    mark(failed-13, that-10)
    nsubj(failed-13, I-11)
    aux(failed-13, had-12)
    ccomp(certain-9, failed-13)
```

Figure 3.1: Dependency parser output sample.

- *Polarity Detection.* Once the four elements (Subject, Object, Action, and Event) have been identified from a sentence, we need to determine the polarity of each of them respectively. The polarity word associated with each element can be determined from the dependency parse results. We focus on classifying each sentence element as positive, negative or neutral, instead of combining all the lexicons, we perform polarity detection separately using each of the lexicons and determine the final polarity by majority vote as explained above.
- *Sentence Tense Detection.* To detect the tense of a sentence, we rely on the POS tagging results of verbs for the identification of present or past tense see 3.2. For future tense detection, we use some simple tag patterns (shown in 3.2) along with the verb sense to identify frames associated with that sense in the FrameNet<sup>5</sup>. If the frame is associated with “desiring”, the tense of the verb is identified as future tense.

<sup>5</sup><https://framenet.icsi.berkeley.edu/fndrupal/>

Table 3.2: Tense Patterns

<b>Tense</b>	<b>Pattern</b>
Present	"VBG", "VBP", "VBZ", "VB"
Past	"VBD", "VBN"
Future	("going to have been", "VBG") ("will have been", "MD") ("going to have", "VBG") ("look forward to", "VBP") ("will have", "MD") ("will be", "MD") ("going to", "VBG") ("hope to", "VBP") ("will", "MD")

### 3.2.2 Polarity Detection

Here, we explain various methods and algorithms used to carry out some of the NLP Tasks used throughout this thesis.

#### 3.2.2.1 Keyword Detection

Our keyword detection approach is simple and direct. It basically helps us answer the following question: if given a word, is there an occurrence of the said word in a collection of words or lexicon. First, we read the list of words from a file into an array, then we filter the array removing all words that do not start with the first letter of said word, then we iterate through the list to find and extract the string match of the word. Then a True or False output is returned depending on if the word was found in the list or not. We use this approach for determining if a word is an Intensifier, a Negation word or a stopword.

#### 3.2.2.2 Negation Handling

Within a sentence, while some words are clearly positive or negative, there are words whose valence can be altered by other words close by. There are various works that have shown that proper detection of negation and intensifiers improves the accuracy of sentiment or emotion detection [287, 206, 179, 265]. Overall, dealing with negation involves two subtasks, the first task deals with finding the words in a sentence which signal negation (negation words). The second task identifies the words within the sentence context that are affected (negated) by the identified negation word. There are various approaches for negation detection and handling of the scope of negation.

Li et al [136], approached the problem as a simple shallow semantic parsing problem for negation scope finding, using the negation word as a predicate for mapping the constituents arguments which make up the negation scope of the negation word. Hogenboom et al [103] in their work considered various negation scope determination approaches which included negating the rest of the sentence after negation keyword, negating the first sentiment word following the negation keyword, negating the first next non-adverb word following a negation keyword, using a fixed context window before or after a negation keyword. Vilares et al [276] in their work used dependency types to determine negation and the scope of the negation. the accuracy obtained by this approach is related to the parsing accuracy.

We create a list of about 110 English intensifiers (see Appendix C) and about 55 English negation words. In our experiments, we worked with a context window of 2, which is 2 words before the keyword. we choose a context window of 2 as various researchers have shown that it produces reasonably good results when compared to higher window sizes for negation handling[103, 59]. The semantic intensity of the word can be modified by an intensifier, and this semantic intensity can be increased by intensifiers called boosters or decreased by intensifiers called down-toner. Thelwall et al in their work, used a booster word list that increases or reduces emotion strength by 1 or 2 [265]. However, identifying the specific weights of intensifiers is an area of research, we use a more simplified weight approach: if the default weight of every word is 1, we assume that booster or down-toner should increase or decrease that weight by 50%, thus we assign a weight or score to the intensifiers: 0.5 for boosters and -0.5 for down-toners. We identify the presence of intensifiers within the context window and add the weight to the score of the keyword. We identified 3 possible frequently used negation patterns which include:

1. *Negation + Intensifier + Keyword*

For example: I *don't* really **like** you.

In this example, the keyword is “like”, and the context window are the 2 words before it. The negation word is “don’t” and the intensifier is “really”.

2. *Word + Negation + Keyword*

For example: I really *don't* **like** you.

In this example, the keyword is “like”, and the context window are the 2 words before it. The negation word is “don’t” and the intensifier is “really”.

### 3. *Negation + Word + Keyword*

For example: This is not a **problem** for you.

In this example, the keyword is “problem”, and the context window are the 2 words before it. The negation word is “not”. This pattern we found not to be as frequently used as the other two patterns.

The *GetContextString* Function is a simple sub-string retrieval function that gets the first 2 words (if any) before the specified word in a given string (sentence). The *GetWordBefore* Function is a simple sub-string retrieval function that gets the first word (if any) before the specified word in a given string (sentence). Once we have determined that a given keyword is negated, we switch the original polarity of the word from Negative to Positive and vice versa. However, if the keyword is an emotion-bearing word, we use WordNet to lookup the antonyms of said keyword, and apply the emotion of the first antonyms found in the emotion lexicon.

For example: I really *don’t* **love** you.

In this example the keyword is “love”, since it is negated the antonym lookup will return “hate” and hate becomes the emotions attached to the sentence.

---

**Algorithm 1:** Context Valence Detection(Negation Handling)

---

**input** : Sentence, keyword, ContextSize

**output:** bool IsNegated: True or False

isnegated=false;

contextText = GetContextString (*sentence,keyword,windowSize*);

**if** contextText IsNotNullOrEmpty **then**

    beforeWord = GetWordBefore (*contextText,keyword*);

**if** IsIntensifier (*beforeWord*) **then**

        beforeWord = GetWordBefore (*contextText,beforeWord*);

**if** IsNegation (*beforeWord*) **then**

            isnegated=true;

**end**

**else**

**if** IsNegation (*beforeWord*) **then**

            isnegated=true;

**else**

            beforeWord = GetWordBefore (*contextText,beforeWord*);

**if** IsNegation (*beforeWord*) **then**

                isnegated=true;

**end**

**end**

**end**

**end**

return isnegated;

---

### 3.2.2.3 Polarity/Sentiment Detection based on Majority Voting

Here, we use a simple lexicon-based detection system to identify polarity or sentiment using 5 lexicons. We choose a lexicon based implementation for its simplicity in implementation and also its relatively good performance in various research. In order to detect the polarity of words and sentences, we required the use of lexical resources that contain a list of opinion-bearing words. These

words in some resources contain additional meta-data such as a strength score, part of speech tags, and many more. Some lexical resources may contain a list of just positive and negative words, but with syntactic information and there are lexical resources which contain semantic information such as senses of the positive and negative words. We wanted to have a system that considers these 3 lexical resource types. To this end, we used the following lexicons:

1. SentiWordNet [70]: This is an opinion mining resource that allocates sentiment score to synsets in the WordNet. Each synset is assigned as score for objectivity, positivity, and negativity. It falls under the category of lexical resources which contain semantic information such as senses, and has been successfully used in various research across multiple domains. We multiply the score by +1 for each positive terms and -1 for each negative terms.
2. Bing Liu Sentiment lexicon [140]: This lexicon contains about 6789 terms of both negative and positive items. The list holds no scores, no semantic or syntactic information. The lexicon also contains words that are not spelt correctly, as it aims to try and represent the use of language on the Internet. We assign +1 to each positive terms and -1 to each negative terms.
3. Subjectivity lexicon [287]: Each word in this lexicon has a subjectivity level, part of speech tag, stemmed/unstemmed, and semantic orientation. The lexicon provides cases where words have a different orientation based on its part of speech tag. it also contains words as well as phrases totaling 8225 subjectivity words. We assign +2 to each strongly subjective positive terms and +1 to weak subjective positive terms. We assigned -2 to each strongly subjective negative terms and -1 to weak subjective negative terms.
4. AFINN lexicon [185]: is a list of positive and negative English words with polarity scores between -5 and +5. The words in this list have been human annotated with about 2477 entries of both positive and negative words. We made use of the newer version of this lexicon.
5. Q-WordNet [3]: This is an automatically generated lexicon from the Senti-WordNet. It associates sense to a quality of a positive or a negative. The experiments conducted in various research show that it preforms better than Senti-WordNet [3, 275].

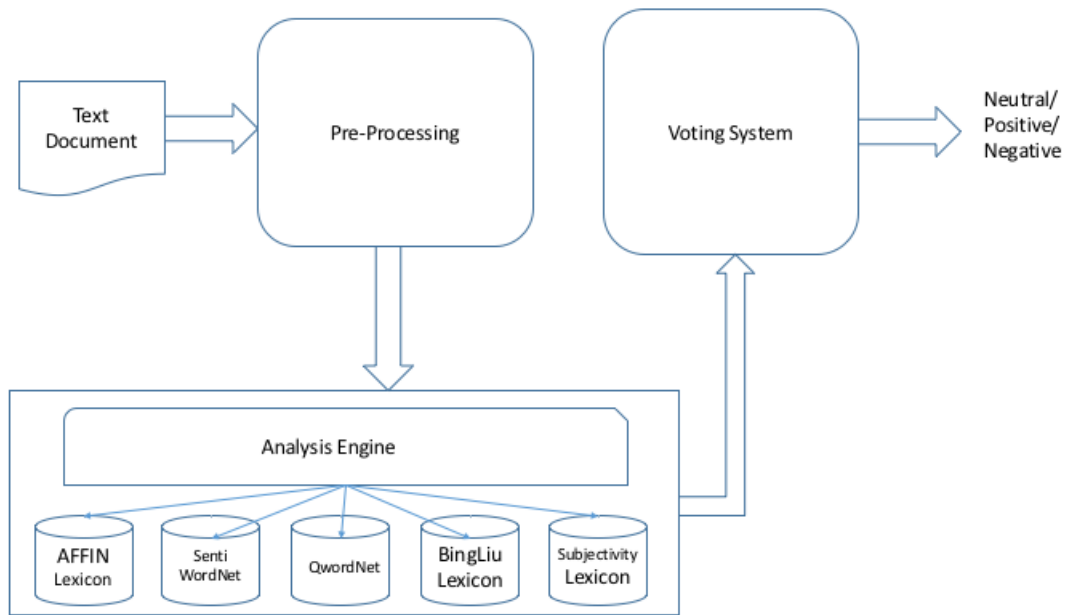


Figure 3.2: Polarity/Sentiment Detection based on Majority Voting

We first perform pre-processing on text such as *Sentence splitting and tokenisation*<sup>6</sup>, *Part-of-Speech Tagging*, *Dependency Parsing*<sup>7</sup>. Note that we did not remove stop words as it would become difficult to identify proper word senses without the presence of stop words. For example, using the list of stop words from Stanford NLP (CoreNLP)<sup>8</sup> the sample sentence below becomes

example with stopwords

```
{
  "text": "I would like to go on holiday not work",
  "partOfspeechTags": "PRP MD VB TO VB IN NN RB VB"
}
```

example without stopwords

```
{
  "text": "like go holiday work",
  "partOfspeechTags": "IN JJ NN NN",
}
```

<sup>6</sup><http://nlp.stanford.edu/software/tokenizer.shtml>

<sup>7</sup><http://nlp.stanford.edu/software/stanford-dependencies.shtml>

<sup>8</sup><https://github.com/stanfordnlp/CoreNLP/blob/master/data/edu/stanford/nlp/patterns/surface/stopwords.txt>



The example shows that the POS tags of various words have changed, which will in turn affect the sense of the word. For each sentence in a document, we carry out lexicon-based with each of the lexicons individually using the algorithm below. Our voting system has three candidates Positive, Negative, and Neutral. Each of the 5 lexicon implementations will return Positive, Negative or Neutral which counts as a vote for the corresponding candidate for a particular sentence. The sentence/phrase or word is assigned the value of the candidate with the highest vote.

---

**Algorithm 2:** Sentence level Lexicon based calculation

---

**input** : Sentence

**output:** Positive or Negative or Neutral

SentenceScore=0;

contextWindow=2;

Tokens = Tokenize(Sentence);

Tags = POSTagger(Sentence);

**for**  $i = 0; i < Tokens.Count; i = i + +$  **do**

    currentSentenceToken = Tokens[i];

**if** *LexiconList Contains currentSentenceToken* **then**

        LexiconItem = GetLexiconItem(currentSentenceToken);

        LexiconItem = LexiconItem.Value +

        GetIntensifierScore(currentSentenceToken,Sentence,contextWindow);

**if** *HasNegation(currentSentenceToken,Sentence,contextWindow)* **then**

            Reverse Lexicon term value by multiplying by -1 so +ve becomes -ve and vice

            vasa

**end**

**end**

    SentenceScore= SentenceScore + LexiconItem.Value;

**end**

SentencePolarity = Neutral;

**if** *SentenceScore > 0* **then**

    SentencePolarity = Positive;

**end**

**if** *SentenceScore < 0* **then**

    SentencePolarity = Negative;

**else**

**end**

return SentencePolarity;

---

### 3.2.3 Assigning Values to the OCC Variables

We now describe how we assign values to each OCC variable.

**Direction:** The value for this variable can either be “Self” or “Other”. The former refers to emotions expressed for oneself, while the latter refers to emotions expressed for others. This value is assigned based on the dependency relationship (identified by the dependency parser) of a first-person pronoun (such as “I”, “me”, “we”) with an action or event. We identify 3 possible scenarios for assigning a value to this variable:

1. When dealing with a simple sentence, we simply apply the process mentioned above
2. When dealing with a complex sentence where multiple subjects are identified by the parser, we assign values based on respective action/event relations with identified subjects;
3. No subject is identified OR no verbs exist in the text, here we just assign the value “Other” to the variable.

**Tense:** The value for this variable can either be “Present”, “Past” or “Future”. The value assignment is determined by the POS tags of the verbs in a sentence or by the results obtained from the FrameNet as has been described in section 3.2.1. In the cases where no verbs are used in a sentence, the value of the variable is set to “Present”.

**Overall Sentence Polarity:** The value for this variable can either be “Neutral”, “Negative” or “Positive”. This is the overall polarity of a sentence which is determined by the polarity detection method.

**Event Polarity:** The event is identified based on the verb-object relations revealed by the dependency parser. The noun phrase which contains an identified object is treated as the event for its relative verb. The polarity of an event is then determined using the aforementioned polarity detection method.

**Action Polarity:** The action is identified based on the subject-verb relations revealed by the dependency parser. The verb phrase which contains the identified verb is treated as an action. Similar to event polarity, the action polarity is also determined using the aforementioned polarity detection method.

We demonstrate with a walk-through sample given the sentence “*When I passed the first examination that I had to repeat.*”.

- **Direction:** The subject of the sentence is the pronoun “*I*” identified by the dependency parsing result “*nsubj(passed-3, I-2), nsubj(had-9, I-8)*” and assigned the value **SELF**.
- **Tense:** This is identified by the verb tag “*passed/VBD*”, which is associated directly with the subject, object and event based on the dependency results “*nsubj(passed-3, I-2), nsubj(had-9, I-8), dobj(passed-3, examination-6)*”. The Tense variable is and assigned the value **PAST**.
- **Overall Polarity:** the overall polarity of the sentence is **POSITIVE**.
- **Event Polarity:** The event, which is the verb + object combination “*passed examination*” has a **POSITIVE** polarity.

**If** Direction = “Self” **and** Tense = “Past” **and** Overall Polarity = “Positive” **and** Event Polarity = “Positive”, **then** Emotion = “Satisfaction”.

walk-through sample

```
{  
  
  "ExpectedEmotion": "Joy",  
  "Sentence": "When I passed the first examination that I had to repeat.",  
  "DataSet": "ISEAR",  
  "Analysis": [  
    {  
      "PosTagging": "When/WRB I/PRP passed/VBD the/DT first/JJ examination/N",  
      "ContainsNegation": false,  
      "OverallPolarity": "Positive",  
      "Tense": "Past",  
      "Elements": [  
        {  
          "Type": "Object",
```

```

    "Phrase": "examination",
    "polarity": "Positive"
  },
  {
    "Type": "Event",
    "Phrase": "passed examination",
    "polarity": "Positive"
  },
  {
    "Type": "Subject",
    "Phrase": "I",
    "Direction": "Self",
    "polarity": "Positive"
  }
],

"Dependency": [
  "advmod(passed-3, When-1)",
  "nsubj(passed-3, I-2)",
  "root(ROOT-0, passed-3)",
  "det(examination-6, the-4)",
  "amod(examination-6, first-5)",
  "dobj(passed-3, examination-6)",
  "mark(had-9, that-7)",
  "nsubj(had-9, I-8)",
  "ccomp(passed-3, had-9)",
  "aux(repeat-11, to-10)",
  "xcomp(had-9, repeat-11)"
],

```

```

"Identified_Emotions": [
  {
    "Emotion": "Satisfaction",
    "EmotionMap": "Joy",
    "Desc": "pleased about the confirmation of a desirable event"
  }
]
}
]
}

```

Table 3.3: Variable assignment and possible values.

Variable	Possible Values	Description
Direction	Self or Other	The value is determined by the dependency relationship of first-person pronoun with an action or event.
Tense	Present, Past or Future.	Determined by using POS tags of the verbs in the sentence
Overall Polarity	Neutral, Negative or Positive	Overall polarity of a sentence which is determined by the polarity detection method aforementioned
Event Polarity	Neutral, Negative or Positive	The event is the verb-object relations revealed by the dependency parser. The noun phrase which contains an identified object relative to the verb is treated as the event
Action Polarity	Neutral, Negative or Positive	The action is identified based on the subject-verb relations revealed by the dependency parser. The verb phrase which contains the identified verb is treated as an action.

Once the variable values are identified, the rules defined in Table 3.1 are then applied to detect the presence of emotions. The compound emotions are results of the output of the event-based and

action-based emotions. For the "sorry-for" emotion, we ensure that the subject is of positive valence; otherwise, the emotion is identified as "resentment". The same rule is applied to the "admiration" and "reproach" emotion pairs.

### 3.3 Experiments

In this section, we present the evaluation results of our OCC-based emotion detection approach on three different datasets, which include: *The International Survey On Emotion Antecedents And Reactions (ISEAR) Dataset*, *The SemEval-2007 Task 14 Affective Text dataset* [249] and *The Alm's Dataset* [6].

- *The International Survey On Emotion Antecedents And Reactions (ISEAR) Dataset* was collected during the 1990s by a large group of psychologists. This dataset contains 7,667 sentences labelled with seven emotions (*joy, fear, anger, sadness, disgust, shame, guilt*) which was developed by asking nearly 3,000 people with different cultural background about various emotional events they have had.
- *The SemEval-2007 Task 14 Affective Text dataset* [249] consists of news headlines collected from major newspapers such as New York Times, CNN, and BBC News. The dataset had a training set of 250 annotated headlines, and a test set with 1,000 headlines. Each news headline is labelled with six emotions (*joy, fear, anger, sadness, disgust*). In our work here, we use only the test set.
- *The Alm's Dataset* [6] contains 1,207 sentences annotated with five emotions (*sad, angry-disgusted, happy, fearful*) extracted and annotated from 176 fairy tale stories. In our experiments, we use only the data extracted from Grimm's and Anderson's tales, which have a total number of 1,040 sentences.

As our goal is to detect emotions in the absence of emotion-bearing words, we filter out sentences which contain emotion words as can be found in the emotion lexicon WordNet-Affect<sup>9</sup>. The total number of sentences before and after filtering of emotion-bearing words in each emotion category for these three datasets are shown in Table 3.4. It can be observed that according to WordNet-Affect, 45% sentences in ISEAR and 87% news headlines in SemEval do not contain any

---

<sup>9</sup><http://wndomains.fbk.eu/wnaffect.html>

emotion words. Thus, it is quite common for people expressing emotions without the explicit use of emotion words. Only 35% of sentences in fairy tales (the Alm's dataset) do not contain emotion words. This is expected since when telling stories to children, the good and bad lines are usually made very clear. Thus, one can say that sentences derived from fairy tales dataset often tend to contain explicit expressions of emotions.

It is worth noting that while each sentence in ISEAR or Alm's was only labeled with one emotion label, each news headline in the SemEval dataset can be labeled with multiple emotions, and each emotion was further annotated with a score in [0, 100] indicating the degree of emotion load. We take the emotion with the maximum score as the label for each news headline<sup>10</sup>.

We focus specifically on the 5 emotion categories which are shared across these three datasets and map the OCC-output emotions to the five emotion categories in the following ways:

- Fear: Fear, Fear-confirmed
- Joy: Joy, Happy-For, Satisfaction, Admiration, Pride
- Anger: Anger, Reproach
- Sadness: Distress, Sorry-For, Disappointment, Shame
- Disgust: Resentment

We have also developed three baseline models. One is a lexicon matching method which uses the NRC emotion Lexicon<sup>11</sup> for sentence-level emotion detection, We also train supervised Naïve Bayes (NB), Support Vector Machine (SVM) classifiers using the implementation in Weka<sup>12</sup> on the three datasets with 10-fold cross-validation. NB and SVM were trained using unigram, bigram and trigram for words. We report the results in terms of the F-measure scores, which is commonly used to report such experiments along with recall and precision scores. Below are our experimental results and findings for all and each individual dataset.

---

<sup>10</sup>In [249], it was suggested that for coarse-grained evaluation on SemEval, each emotion should be mapped to a 0/1 classification (0 = [0,50), 1 = [50,100]). However, we found that only 342 out of a total of 1000 news headlines are left with emotion labels using this mapping method which would be insufficient for our experiments.

<sup>11</sup><http://www.saifmohammad.com/WebPages/lexicons.html>

<sup>12</sup><http://www.cs.waikato.ac.nz/ml/weka>



Table 3.4: Statistics of the three datasets used in our experiments. “Total” denotes the original number of sentences in each emotion category while “Implicit” denotes the number of sentences which do not contain any emotion words according to WordNet-Affect.

Emotion	Total	Implicit		Emotion	Total	Implicit	
Joy	1095	537	<b>(a) ISEAR</b>	Joy	362	317	<b>(b)</b>
Fear	1095	366		Fear	160	130	
Anger	1096	483		Anger	66	60	
Sadness	1096	488		Sadness	202	182	
Disgust	1096	484		Disgust	26	24	
Shame	1096	581		Surprise	184	160	
Guilt	1093	482		Total	1000	873	
Total	7667	3421					

Emotion	Total	Implicit	
Happy	406	103	<b>(c)</b>
Fearful	121	33	
Angry-Disgusted	174	84	
Sad	247	90	
Surprised	92	50	
Total	1040	360	

**SemEval**

### Alm’s

#### 3.3.1 Results and Analysis

It can be observed from Table 3.5 that although we have filtered out sentences which contain emotion words from WordNet-Affect, using other emotion lexicons such as the NRC emotion lexicon can still identify emotions of some sentences. The simple lexicon matching method has very low F-measure values and on average only achieves 33.35% in F-Measure for all emotion categories across all 3 datasets. This is not surprising since most sentences do not contain any emotion-bearing words. It fails to identify any sentences expressing the “Fear” emotion. Surprisingly, our unsupervised OCC-based approach outperforms supervised NB in three emotion categories “Joy”, “Anger” and “Sadness”. Its overall average F-measure of 53% improving upon lexicon matching by about 20% and better than NB across all three datasets by 10% and SVM by about 6%. If excluding the worst performing “Fear” category, our approach even outperforms NB nearly 6% in F-measure.

For the SemEval dataset 3.5, our approach performs best on the “Sadness” category with 62% F-measure. The worst performance is still in the “Fear” category (31% in F-measure). For the rest three emotion categories, our approach achieves an average F-measure of around 60%. Compared

Table 3.5: Performance comparison of F-measure (F) results on the 3 datasets. Alm dataset(F1), ISEAR(F2) and SemEval(F3).

Emotion	NRC Lexicon			Supervised NB			Supervised SVM			Our Approach		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Joy/Happy	58.76	33.42	39.68	56.10	49.60	56.60	58.00	62.40	60.50	61.67	69.55	59.16
Fear/Fearful	0	0	0	32.70	46.80	38.30	19.60	52.40	54.80	14.04	18.27	31.79
Anger/Angry-Disgusted	48.92	23.01	55.78	56.60	31.40	16.10	54.60	38.30	27.80	66.57	61.34	61.41
Sadness/Sad	60.98	25.63	47.75	57.60	44.40	34.00	53.80	52.50	46.60	69.50	67.97	62.51
Disgust	-	25.58	38.52	-	39.70	31.30	-	43.50	29.40	-	39.20	61.72
Average	42.17	21.53	36.35	50.75	42.38	35.26	46.50	49.82	43.82	52.95	51.27	55.32
Average (- Fear)	56.22	26.91	45.43	56.76	41.27	34.50	55.47	49.18	41.08	65.91	59.52	61.20

to the NRC lexicon results, our approach gives a superior performance with the average F-measure result improved by 19%. Supervised NB only outperforms our approach on the “Fear” category by about 6%. While SVM outperforms our approach in the “Joy” and “Fear” categories by 1% and 23% respectively. SVM had the best f-measure in the “Fear” category for this dataset, with a score of 54%. In the SemEval dataset, headlines may be labelled with multiple emotions with varying intensity. We have also evaluated the ability of our approach to detect multiple emotions in a sentence and identify instances of full match (identifying all the emotion labels correctly) and partial match (identifying part of the emotion labels correctly). We found that in Figure 3.3, our approach achieves an accuracy of 18% for full match and 53% for partial match. Thus, our approach can indeed detect multiple emotions in sentences.

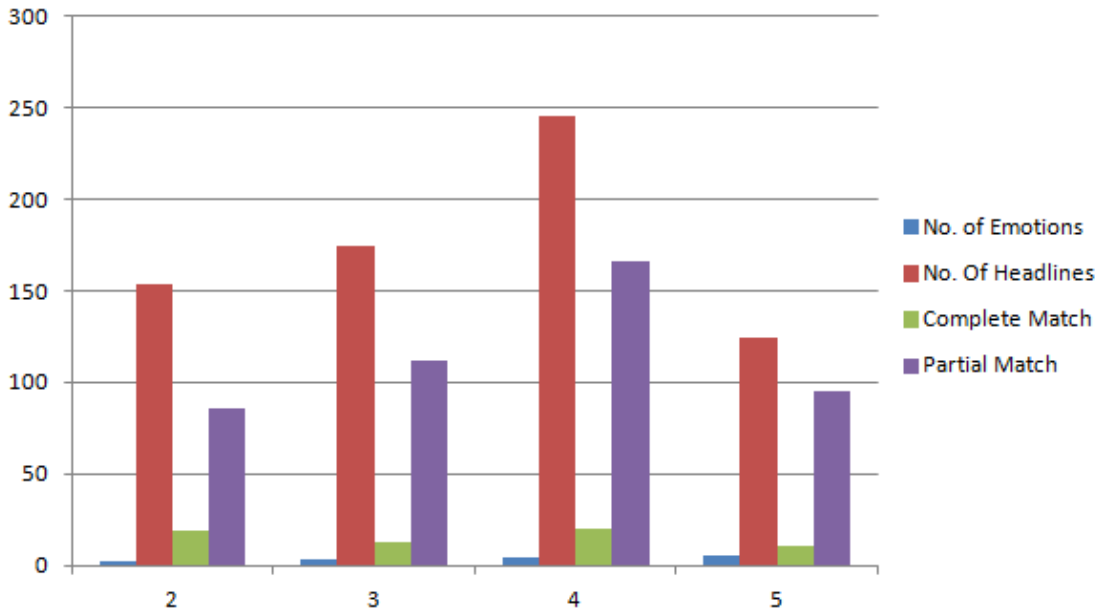


Figure 3.3: Multi-emotion detection distribution using SemEval dataset

For the Alm’s dataset, our approach gives very good results on detecting emotions of “Happy”, “Angry-Disgusted” and “Sad” with the F-measure scores ranging between nearly 61% and over 69%. The improvement over the NRC lexicon labelling baseline is at the range of 19-38%. The worst performance is still on the “Fearful” category where the F-measure score is only 14%. Supervised NB only outperforms our approach on the “Fearful” category and gives worse results on all the other emotion categories. Overall, our approach improves upon NB by 4% (with “Fearful”) and 11% (without “Fearful”) in F-measure, and improves on SVM by 8% (with “Fearful”) and 12% (without “Fearful”) in F-measure.

For the ISEAR dataset, our approach performs best on the “Happy” category with 69% F-measure. “Fear” category is still the worst performing (18% in F-measure). For the rest three emotion categories, our approach achieves an average F-measure of around 63%. Our approach also had a bad performance on the “Disgust” category (39% in F-measure). Compared to the NRC lexicon results, our approach gives a superior performance with the average F-measure result improved by 30%. Supervised NB only outperforms our approach on the “Fear” category by about 27%. While SVM outperforms our approach in the “Fear” category by 37% respectively. SVM had the best f-measure in the “Fear” category for this dataset, with a score of 52%.

### 3.3.2 Error Analysis

Error analysis was conducted on the emotion detection results in order to better understand how the proposed approach performs. Table 3.6 shows some example detection results. Within each step in our approach, we log instances where an error may have occurred or a failure to proceed to the next step is detected such as if returned emotion is neutral, or if WSD returns no results or polarity detection is neutral or misclassification of emotion. In general, whenever the output of one process is the input of another process we track it and log instances where a failure or error may have occurred. We collected a sample size of 16 records for each dataset where an error has occurred and examined them to identify causes of errors and possible mitigation. In total, we reviewed a total of 48 error instances across all 3 datasets.

Looking at example 1 in Table 3.6, our approach detected the comment as containing 2 core sentences *Findings: Can humanity survive?* and *Want to bet on it?* and found both sentences to be in the present tense and hence failed to detect any emotion since fear according to our model, is

Table 3.6: Detection examples of correct and failed detection using our approach on SemEval dataset

<b>Example</b>	<b>Expected Emotion</b>	<b>Detected Emotion</b>
1. <i>Findings: Can humanity survive? Want to bet on it?</i>	Fear	Distress
2. <i>Firms on alert for letter bombs</i>	Fear	Fear
3. <i>Joseph Wambaugh's new start</i>	Joy	Neutral
4. <i>Amazon.com has 'best ever' sales</i>	Joy	Joy
5. <i>That'll cost ya</i>	Anger	Neutral
6. <i>Anglicans rebuke U.S. branch on same-sex unions</i>	Anger	Anger

found on past and future tenses only. Improving on our tense detection approach which we present later in this chapter, can help with mitigating this error.

If we take a look at example 3 in Table 3.6, our approach failed to detect a tense and found the polarity of the sentence and event to be **neutral**. This according to our model, means that it is not a valanced reaction and as such will produce no emotion.

Looking at example 5 in Table 3.6, our approach correctly detected the comment as in future tense. However, our approach found the polarity of the sentence and event "cost ya" to be **neutral**. This according to our model, means that it is not a valanced reaction and as such will produce no emotion. To mitigate against this we would consider adding a fallback hybrid approach for polarity detection to assist when our highest vote approach fails to detect polarity.

In summary, in this dataset our approach failed for 3 major reasons: *Inability to identify the sentences as valanced (polarity)* accounted for 23.7% of overall errors, the low coverage of *rules for specific emotions* accounted for 36.8% of overall errors and reliance on *Verb+Object relationship* for event detection accounted for 12% of the overall errors in this dataset.

Table 3.7: Detection examples of correct and failed detection using our approach on ISEAR dataset

<b>Example</b>	<b>Expected Emotion</b>	<b>Detected Emotion</b>
1. <i>Passing an exam I did not expect to pass.</i>	Happy	Happy
2. <i>When I got the loan for my studies.</i>	Happy	Neutral
3. <i>When I found out that my father had lung cancer and they did not know how long he would live.</i>	Fear	Sorry-For

Looking at example 2 in Table 3.7, our approach failed to correctly detect the emotion here, because it failed to detect the polarity of the sentence and event. According to our model, it is not a

valanced reaction and as such will produce no emotion. To mitigate against this we would consider a hybrid approach for polarity detection to assist when our highest vote approach fails to detect polarity.

Looking at example 3 in Table 3.7, our approach failed to correctly detect the emotion here, because of the rules to detect the specific expected emotions. In example 3, there are 3 subjects identified by the dependency parser *nsubj(had-8, father-7)*, *nsubj(live-20, he-18)* and *nsubj(found-3, I-2)*. However, we found on object associated with the subject 'I', which would lead our system to assign a direction variable of 'other' and in turn categorize the sentence as **Sorry-For**. To mitigate against this type of errors, we need to add new emotion specific rules which are closely related to real world scenarios of specific expressions of emotions.

example 3 explained

```
{
  "Subjects": [
    {
      "word": "I",
      "index": 1,
      "dependencytag": "nsubj(found-3, I-2)",
      "Polarity": 0
    },
    {
      "word": "father",
      "index": 6,
      "dependencytag": "nsubj(had-8, father-7)",
      "Polarity": 0
    },
    {
      "word": "he",
      "index": 17,
      "dependencytag": "nsubj(live-20, he-18)",
```

```

        "Polarity": 0
    }
],
"Objects": [
    {
        "word": "cancer",
        "index": 9,
        "dependencytag": "dobj(had-8, cancer-10)",
        "Polarity": -1
    }
],
"Events": [
    {
        "phrase": "had cancer",
        "Polarity": -1
    }
]
}

```

In this dataset our approach failed for 2 major reasons: 27% of the overall errors were as a result of *Inability to identify the sentences as valanced (polarity)* and *the rules for identifying specific emotion types* and accounted for about 45% of the overall errors

If we take a look at example 1 and 2 in Table 3.8, we find that the emotion of fear is described in example 1 and we found that in this dataset, our approach struggled greatly with such complex sentences. In these cases, our approach failed to identify the emotion as fear, because the direction of the emotion was *other* and hence identified it as *sorry-for*. To mitigate against this type of errors, we need to add new individual emotion-specific rules which are closely related to real-world scenarios of specific expressions of emotions.

Looking at example 3 and 4 in Table 3.8, our approach failed to correctly detect the emotion here, because it failed to detect the polarity of the sentence and event, which it found to be **neutral**.

Table 3.8: Detection examples of correct and failed detection using our approach on Alm dataset

<b>Example</b>	<b>Expected Emotion</b>	<b>Detected Emotion</b>
1. <i>Then she threw herself upon Heaven for help in her need, and went away, and journeyed on the whole night, till at last she came to a large wood.</i>	Fear	Sorry-for
2. <i>The second did not want to go in at all, but was forced</i>	Fear	Distress
3. <i>I don't want to sit here any longer.</i>	Anger-disgusted	Neutral
4. <i>Away with him!</i>	Anger-disgusted	Neutral
5. <i>It is overpowering, and yet it is delightful.</i>	Happy	Happy

This according to our model, means that it is not a valanced reaction and as such will produce no emotion. To mitigate against this we would consider a hybrid approach for polarity detection to assist when our highest vote approach fails to detect polarity.

In this dataset our approach failed for 4 major reasons: *Inability to identify the sentences as valanced (polarity)* which accounted for 17.7% of the overall errors, *the low coverage of rules for specific emotions* which accounted for 22% of the overall errors, *the use of complex descriptive sentences* which accounted for 24.7% of the overall errors, and not paying better attention to the *weight of punctuation marks* such as ? and !, as they seemed to have a major influence in meaning in this dataset and accounted for about 25% of the overall errors encountered.

### 3.3.3 Discussion

It can be observed from Table 3.5 that although we have filtered out sentences which contain emotion words from WordNet-Affect, using other emotion lexicons such as the NRC emotion lexicon can still identify emotions of some sentences, and that using the NRC lexicon fails to detect any “Fear” emotion bearing sentences for all the datasets experimented here. Nevertheless, our approach is still able to identify some of the sentences relating to “Fear”. Despite using no labelled data, our approach achieves similar performance as supervised NB on the ISEAR and SemEval datasets (with 1% difference in F-measure) and outperforms NB by 9.5% in F-measure on the Alm’s dataset. The results also show that our approach is largely affected by the quality of the text. ISEAR contains personal experience expressed by a wide range of participants and hence, might contain lots of informal and ill-grammatical text. SemEval contains news headlines which are often incomplete

sentences ignoring grammar conventions. The Alm's dataset, on the other hand, contains fairy tales which are formal text following rules of grammar very strictly. As such, the performance obtained from the Alm's dataset by our approach is significantly better than that obtained from the other two datasets.

Our approach relies on results generated from a series of NLP tasks such as word-sense disambiguation, POS tagging, dependency parsing, and polarity detection to be able to assign values to a set of OCC variables for emotion detection. Thus, any error that occurs will be propagated down the pipeline process. Furthermore, failure in detecting the polarity of text will make it impossible for our approach to identify the underlying emotion. Also, we have not considered ironic and sarcastic sentences in our current work. Nevertheless, we have shown that in the absence of annotated data, the OCC-based approach is able to identify implicit emotion in text with performance competing with supervised classifiers, and even outperforms the supervised approach for formal text (the Alm's dataset). The emotion detection results generated by the OCC-based approach can be used as seed examples to bootstrap more complicated emotion detection methods which require a large amount of training data.

### **3.4 Improving Rule-Based Approach to Implicit Emotion Detection in Text**

In this chapter, we have introduced our rule-based approach to implicit emotion detection, and have identified some problems of our approach which contributed errors to our system and in turn affected the overall performance of the system. In this section, we explore various ways to address the problems of our initial approach by making improvements to tense detection, sentence type detection and improvements to some of our variable rules.

#### **3.4.1 Tense Detection**

In English grammar, tense can be described as a term that provides a time reference in relation to a moment expressed in writing/speaking [36, 71, 44]. Various verb forms are used to manifest tenses, these verbs often change forms depending on combinations and number of verbs within the sentence[36]. There are basic tenses types found in various languages such as future, past and present[36, 71, 44]. The Penn treebank defines tags that can be used to identify both present and past tenses based on the verbs. However, basing tense detection solely on the POS tags is not good enough as most sentences have multiple verbs with various POS tags and the combination of these



verbs put the sentence in a completely different tense from the POS tags.

In other to improve our current implementation of tense detection, we focus on auxiliary verbs and their combinations. Auxiliary verbs are verbs which are used in creating tenses, voices, and moods of verbs [36, 71, 44]. In the English language, the primary auxiliary verbs are be, do, and have. Auxiliary verbs are must often both stand-alone verbs and auxiliary verbs, this means that they can work with other verbs to create a verb phrase [36, 44, 274].

<b>Be</b>	<b>Do</b>	<b>Have</b>
am	does	has
is	do	have
are	did	had
was		having
were		
being		
been		

Figure 3.4: List of auxiliary verbs and their base forms [274]

<b>Modal Auxiliaries [Never Change Form]</b>
can, could, may, might, must, ought to, shall, should, will, would

Figure 3.5: List of modal auxiliaries [274]

Our approach is based on 3 key elements; the number of auxiliary verbs in the sentence, the base form of identified auxiliary verbs and the Penn treebank tense definition of the identified auxiliary verbs. Our approach is based on Pickbourn’s work on the English verb, which introduces the current modern distinction between tenses, mainly for his recognition of the importance of *context* within sentences. Pickbourn’s work is discussed in detail in Robert Binnick’s book (Time and the verb: A guide to tense and aspect) [36].

Penn Treebank tag tense

{

```

"TenseTags": [
  "VBG": "gerund",
  "VB": "present",
  "VBP": "present",
  "VBZ": "present",
  "MD": "present",
  "VBN": "past",
  "VBD": "past"
]
}

```

### 3.4.1.1 Present Tense

The present tense is a tense type which draws attention to action/actions which are on-going or actions which are performed habitually. It presents information about events which are currently happening from the perspective of the speaker. There are 4 types of present tense which include simple present, present progressive, present perfect, and present perfect progressive.

- **Simple present** tense is used to express events/actions commonly referred to as habitual actions which include factual, normal, or regular in occurring actions/events.
- **Present progressive** tense is used to express events/actions which are currently happening now or currently in progress.
- **Present perfect** tense is used to express events/actions that happened at a non-specific time and actions which started in the past but continued to the present.
- **Present perfect progressive** tense is used to express events/actions which most often started in the past and has either recently stopped or has continued to the present.

### 3.4.1.2 Past Tense

The past tense is a tense type which is used to express actions/events which have already occurred. These actions/events are considered to be finite and as such, they should have a start and a stop



Figure 3.6: Present tense auxiliary examples [36]

<b>Tense</b>	<b>No of Auxiliaries</b>	<b>Base Auxiliary</b>	<b>Penn Tree Tag Tense</b>
present simple	1	do	present
present progressive	1	be	present
present perfect simple	1	have	present
present perfect progressive	2	have,be	present , past

Table 3.9: Present tense Rules

point. There are 4 types of present tense which include simple past, past progressive, past perfect, and past perfect progressive.

- **Simple past** tense is used to express events/actions that began and ended at a specific time.
- **Past progressive** tense is used to express events/actions which in the past lasted for a duration of time.
- **Past perfect** tense is used to express events/actions in the past that completed before another past event/action started.
- **Past perfect progressive** tense is used to express past events/actions which were in progress before other events/actions.

<b>Tense</b>	<b>No of Auxiliaries</b>	<b>Base Auxiliary</b>	<b>Penn Tree Tag Tense</b>
past progressive	1	be	past
past simple	1	do	past
past perfect simple	1	have	past
past perfect progressive	2	have, be	past, past

Table 3.10: Past Tense Rules



Figure 3.7: Past tense auxiliary examples [36]

### 3.4.1.3 Future Tense

The future tense is a tense type which is used to express actions/events which have not yet occurred. These actions/events are yet to happen or will happen at some point in the future. There are 4 types of present tense which include simple future, future progressive, future perfect, and future perfect progressive.

- ***Simple future*** tense is used to express actions/events that will occur.
- ***Future progressive*** tense is used to express events/actions which will be ongoing in the future.
- ***Future perfect*** tense is used to express events/actions which will be completed at future time, as well as events/actions which will be completed before another future event/action.
- ***Future perfect progressive*** tense is used to express future events/actions to be finished at a specified time.



Figure 3.8: Future tense auxiliary examples [36]

Tense	No of Auxiliaries	Base Auxiliary	Penn Tree Tag Tense
future simple (going)	3	be,go,to	present, gerund
future simple	1	will	present
future progressive	2	will , be	present, present
future perfect	2	will, have	present, present
future perfect progressive	3	will, have, be	present, present, past

Table 3.11: Future Tense Rules

---

**Algorithm 3:** Auxiliary Verb Tense Detection

---

```

input : Sentence
output: String Tense: Past or Present or Future
tense=Present;
tokens = Tokenizesentence;
posTags =GetPOSTags(tokens);
rules = GetTenseRules(tokens);
IdentifiedAuxList=[ ];
auxCount = 0; for  $i = 0; i < tokens.Count; i = i + +$  do
    token =tokens[i]; tag =posTags[i];
    if IsAuxVerb(token,tag) then
        auxCount++;
        posTense = GetPOSTagTense(tag);
        base =GetAuxBase(token);
        AddToIdentifiedAuxList(base,posTense);
    end
end
rule = FILTER rules FROM IdentifiedAuxList WHERE auxCount EQUAL rule.NoAux
AND rule.Base CONTAINS AuxListItem.Base AND rule.PennTagTense CONTAINS
AuxListItem.TagTense
tense = rule.Tense;
return tense;

```

---

We re-conducted our experiments using our approach to see if the tense detection improvement had any effect on the overall performance of the our approach. We found that in Table 3.12 the changes improves the average F-measure of SemEval dataset by 0.16%, average F-measure of ISEAR dataset by 0.76% and average F-measure of Alm dataset by 1.98%. Across all 3 dataset the average F- measure is improved by 0.97%. The highest improved performance is seen in the Alm dataset while the least improvement is the SemEval dataset. This is because the SemEval dataset is a collection of news headlines that most often have little or no verbs in them. Across all datasets, there is an increased improvement in the F-Measure of the fear emotion. This tells us that our approach new tense detection approach is particularly better than our old approach in detecting past

and future tenses in sentences.

Table 3.12: Performance comparison of F-measure (F) results on the 3 datasets. Alm dataset(F1), ISEAR(F2) and SemEval(F3).

Emotion	<b>Our Approach</b>		
	F1	F2	F3
Joy/Happy	64.83	69.65	59.53
Fear/Fearful	18.25	19.13	31.91
Anger/Angry-Disgusted	66.92	61.86	61.66
Sadness/Sad	69.73	68.56	62.57
Disgust	-	40.95	61.72
Average	54.93	52.03	55.48

### 3.4.2 Using Word Relatedness Measure to Improve Implicit Emotion Detection

In linguistics, words are classified not only by meanings but also by their co-occurrence with other words. To be able to automatically identify semantically associated words is a cornerstone for many NLP applications. This study of co-occurrence of words is based on the assumption is that words that frequently appear together in text are conceptually related. Words which often occur in similar contexts also often tend to have meanings which are similar. This is done mostly to ensure that within a corpus, all related documents to a particular word are considered for said NLP task. This is often referred to as a co-occurrence based word association measure. These word association measures calculate the strength of association between two words by comparing the statistical distribution of the word pairs to the distribution of the individual words in the pair. This is done using the bigram frequency of both words to a function of the individual word unigram frequencies at a corpus-level.

The linguist J. R. Firth in 1957, made a statement which is closely associated with this principle “You shall know a word by the company it keeps” [111]. The distribution of words around a given word is closely related the meaning of a word. In this section, we make use of Pointwise Mutual Information (PMI) and Retrofitted Word Embeddings with Semantic Lexicons in an effort to improve the performance of our implicit emotion detection approach.

#### 3.4.2.1 Extracting Emotion-bearing Words

As previously stated, not all words in a sentence are emotion bearing words, in most cases these words fall into the categories of Nouns, Adjectives, Verbs and Adverbs (NAVA) words. Going by

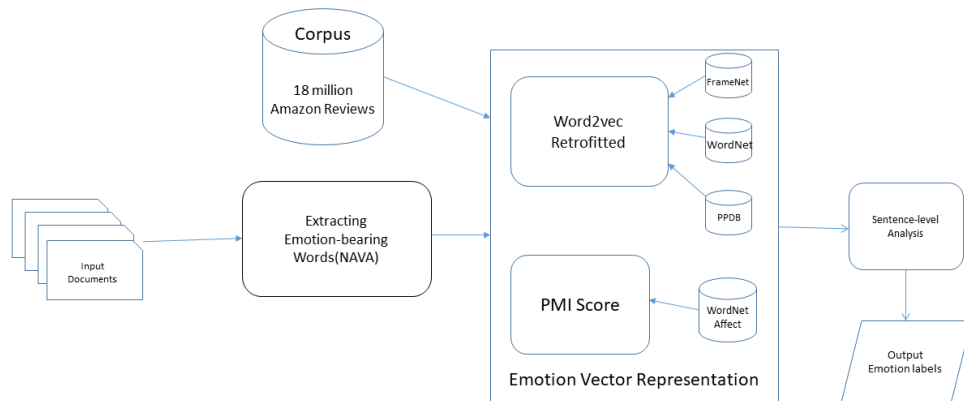


Figure 3.9: Our Approach

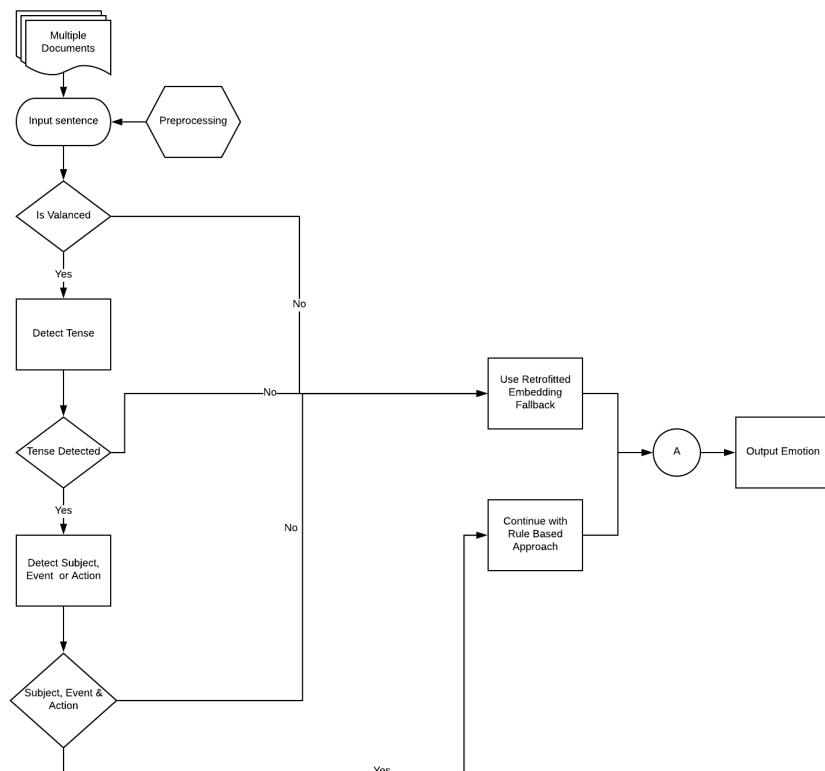


Figure 3.10: Fallback Flow chart

this assumption, we use the part-of-speech tagger to tag, identify and extract these NAVA words. For example, for sentence “*I failed my exams*”, the NAVA words *failed* and *exams* are tagged Verb

(VB) and Noun (NN) respectively. Due to the effect of negation on emotion detection, We check our NAVA words during extraction for negation and replace the negated NAVA word with the WordNet antonym of the word based on its sense. This essential as words convey different meanings/emotions depending on the context.

### 3.4.2.2 Emotion Vector Representation

Vector distributional matrix is based on a co-occurrence and a way of showing the frequency word co-occurrence. However, that simple frequency is not the best measure of association between words. One problem is that raw frequency is very skewed and not very discriminative. When words co-occur more frequently, they are often semantically related. Measuring semantic relatedness is based on the principle that the meaning of a word can be induced by observing its statistical usage across a large sample of language.

Pointwise mutual information (PMI) is a measure of association and in computational linguistics, PMI has been used for finding collocations and associations between words. For instance, counting of occurrences and co-occurrences of words in a text corpus can be used to approximate the probabilities  $p(w)$  and  $p(w,c)$  respectively. Mathematically, the PMI between a word  $w$  and a context word  $c$  is calculated as follows:

$$PMI_{w,c} = \lg \frac{P(w,c)}{P(w)P(c)}$$

where  $\text{occurrence}(w)$  is the number of times that  $w$  appears in a corpus, and  $\text{co-occurrence}(w,c)$  is the number of times that  $w$  and  $c$  co-occur within a specified window in the corpus. The numerator in the equation above provides information given a corpus the frequency two words co-occur. While the denominator provides information assuming each word occurred independently, how often the two words to co-occur. However, PMI tends to be biased toward low frequency co-occurring events and words which are rare in the corpus often have high PMI values. To mitigate against this bias is to slightly change the computation for  $P(c)$  as follows:

$$P_c = \lg \frac{\text{count}(c)^\alpha}{\sum \text{count}(c)^\alpha}$$

Levy et al, Pennington et al and Mikolov et al [134, 161, 202] found that in the equation above



setting  $\alpha = 0.75$  on a variety of tasks yielded good performance. Most research that utilises PMI uses Wikipedia as the corpus for PMI calculation. However, Wikipedia is a resource that is intended to be void of sentiment and emotion as it advocates a neutral point of view. As a result, we have chosen to use Amazon's data corpus of 18 million reviews. We use a window size of 15 words as previous findings report that counting co-occurrences within small windows of text produces better results than larger contexts [43].

Word embeddings are techniques in which words are represented in a predefined vector space with real-valued vectors. The usage words is used to learn this representation. As a result, words which have similar usage often have similar representations and in turn, their meanings are captured in the representation. There are publicly available pre-trained English word vectors, most of which are created from Wikipedia or News corpus. However, these sources are designed to be objective sources of information. We generate word vectors using the word2vec on the Amazon dataset which contains 34,686,770 reviews, vectors were trained on a vocabulary size of 1,507,768 English words and are of length 50.

In other to retrofit our vectors, we use the implementation provided in Faruqui et al [72]. In their work, they indicated that word vectors are a reflection of the data corpus it was trained on and as such, there are words which could make the vectors better and are available outside the data corpus. Thus with their retrofitting approach, additional information can be added to a set of word vectors by adjusting the vectors. They use three different semantic lexicons to retrofit with aim of improving the word vectors.<sup>13</sup>

- **WordNet Synonyms:** WordNet is a handmade semantic lexicon of English words, which puts words into sets of synonyms called synsets, and holds the semantic relations between synsets [162].
- **FrameNet:** FrameNet is a linguistic resource for lexical and predicate argument semantics in English. Each frame hold word types and words that can invoke the same frame are considered to be semantically related [17].
- **The Paraphrase Database:** The paraphrase database is a semantic lexicon containing more than 220 million paraphrase pairs of English. 8 million are lexical single word paraphrases.

---

<sup>13</sup>source code for retrofitting algorithm <https://github.com/mfaruqui/retrofitting>

The key intuition behind the acquisition of its lexical paraphrases is that two words in one language that align, in parallel text, to the same word in a different language, should be synonymous [200].

<b>Before Retrofitting</b>		<b>After Retrofitting</b>	
prostitution	0.7041	theft	0.9220
vandalism	0.7030	larceny	0.9149
under-the-table	0.7012	stealing	0.9071
harassment	0.6919	thieving	0.9010
gambling	0.6903	felony	0.9000

Table 3.13: Top 5 related words returned for *shoplifting* before and after retrofitting

You can see from Table 3.13, that for the word *shoplifting*, the related words returned after retrofitting have better relatedness with the keyword.

We use Ekman's emotion six emotions (happiness, sadness, anger, fear, surprise, disgust) and compute the emotion vector for the individual NAVA words using the semantic relatedness of the words to the emotion category for both word embeddings and PMI. For PMI, we identify the relatedness for each NAVA word to an emotion category by calculating the PMI of said NAVA word to each word in each emotion category as defined in WordNet Affect lexicon. While with Word embeddings, for each NAVA word, we identify the top 20 related words and try to match them with the words in each emotion category as defined in WordNet Affect lexicon.

### 3.4.2.3 Sentence Level Analysis

The overall emotion vector of a sentence is calculated using the arithmetic mean of all identified vector representation for each of the six emotion categories for both PMI and Word embedding respectively. This is computed by aggregating the emotion vectors of all the affect words identified in relation to the NAVA words. The sentence label is derived by choosing the most frequent emotion in the sentence. The emotion category with the highest vector average is assigned to the sentence. If there is a tie, one of the highest emotions is randomly selected.

### 3.4.2.4 Experiment Results

For our experiments, we use the ALM and ISEAR dataset. We re-conducted our experiments using our approach to see if the using the semantic relatedness as a fallback in the event that we cant

detect the correct emotion with our rule-based approach. We found that in Table 3.14 the best performing technique on the Alm dataset is the retrofitted word embedding generated from the 18 million Amazon reviews. Using the Amazon Word embeddings as a fallback for our rule-based approach improves the average F-measure of **Joy** by 1.43%, **Fear** by 26.76% and **Anger** F-measure by 0.28%. The highest improved performance in the Alm dataset is the Fear category. Across all approaches PMI and Word Embeddings, there is a poor performance in F-Measure of the fear emotion. This is because the Alm dataset occasionally uses idioms such as 'make your blood run cold' or 'shake like a leaf' to express fear.

Table 3.14: Tables showing results of the ALM data set

<b>Emotion</b>	<b>PMI</b>	<b>Wiki</b>	<b>Amazon</b>	<b>Amazon + Our Approach</b>
<i>Anger-Disgust</i>	0.36	0.41	<b>0.50</b>	0.67
<i>Joy</i>	0.13	0.24	<b>0.46</b>	0.63
<i>Sadness</i>	0.29	<b>0.41</b>	0.40	0.69
<i>Fear</i>	0.10	0.19	<b>0.32</b>	0.40

We found that in Table 3.15 the best overall performing technique across all emotion categories on the ISEAR dataset is the retrofitted word embedding from Amazon reviews. It improves the F-measure of **Joy** by 0.05%, **Fear** by 13% and **Anger** F-measure by 2% when combined with our rule-based approach as a fallback. The highest improved performance in the ISEAR dataset is the Fear category. Across all approaches PMI and Word Embeddings, there is a poor performance in F-Measure of the fear emotion.

Table 3.15: Tables showing results of the ISEAR data set

<b>Emotion</b>	<b>PMI</b>	<b>Wiki</b>	<b>Amazon</b>	<b>Amazon + Our Approach</b>
<i>Anger-Disgust</i>	<b>0.59</b>	0.55	0.58	0.63
<i>Joy</i>	0.36	0.49	<b>0.51</b>	0.69
<i>Sadness</i>	0.36	0.41	<b>0.48</b>	0.67
<i>Fear</i>	0.20	0.16	<b>0.31</b>	0.44

In this section, results of evaluations show that using semantic relatedness fallback for our rule-based approach yields significantly improved results, especially on the fear emotion category. One of the weaknesses of our approach is that the semantic relatedness scores depend on the text corpus from which they are derived. This means that in order to improve performance we could retrofit

the vectors specifically for each emotion category. From the empirical results, we observed that the retrofitted embeddings perform better than PMI.

In this chapter, we presented our work on implicit emotion detection using the OCC model, with the aim of addressing our first research question: *Can we identify implicit expressions of emotion in text without the use of labeled data?*.

To achieve this, we propose a rule-based approach for implicit emotion detection, using the OCC model. We conducted experiments with three distinct datasets and compared our results with standard baseline supervised classifiers. We showed that in the absence of annotated data, the rule-based OCC model approach is able to identify implicit expressions of emotion.

## Chapter 4

# Detection of Expressions of Blame and Praise in Text

Our investigations and work with the OCC model revealed that there are various emotions such as guilt which require one to establish that blame/praise exists in other to state that such emotions are present. In this chapter, we present our approach for detecting of blame and praise in text documents. We construct a corpus which contains individuals comments relating to their respective on an emotional experience and annotate them for blame and praise.

### 4.1 Introduction

In this chapter, we focus on detecting blameworthiness based on the “Path Model of Blame” presented in the work of Malle et al [145]. In particular, we propose an approach by adapting the original Path Model of Blame and combine some natural language processing techniques for the detection of blames or praises expressed in text. To the best of our knowledge, this is the first piece of work exploring an automated approach for blame/praise detection from text. In order to evaluate our proposed approach, we have created an annotated corpus by labeling each sentence as expressing “*blame*” or “*praise*” from the ISEAR<sup>1</sup> (International Survey On Emotion Antecedents And Reactions) data set. We have also provided annotations at the more finer granularity level to further distinguish the direction of blame/praise, i.e., “*self-blame*”, “*blame-others*”, “*self-praise*” or “*praise-others*”. Our results show that even though our approach gives similar results compared to supervised classifiers on “*blame*”, “*praise*” or “*others*”, it performs better than the supervised classifiers with finer-grained classification of determining the direction of blame and praise, despite using no labeled training data.

Prior to deciding on adapting the “Path Model of Blame” for text-based blame and praise de-

---

<sup>1</sup><http://www.affective-sciences.org/researchmaterial>

tection, we considered other approaches in relation to judgement and the attribution of blame. Such as the work of Mao et al [152], which discusses agent responsibility and attribution. Although their work in most cases is in line with the “Path Model of Blame” in relation to concepts, meaning and definitions, they utilise a more complex model and have a primary focus on the identification of the extraction of the causal reasons as a major path to the attribution of judgement/responsibility. Their approach also depends a lot on manually configured rules and patterns which will not always be present and creates a weakness in the approach.

In the rest of the chapter, we present our proposed approach for blame/praise detection. We explain how we create the annotated dataset and discuss experimental results. Finally, we conclude and outline future directions.

## 4.2 Our Approach

The detection of blame/praise from text can be cast as a classification problem. Where given a comment, we aim to create a model which is able to classify the text as expressing “*blame*”, “*praise*” or “*neither*”. We also explore a finer-grained distinction of the direction of blame and praise, i.e., “*self-blame*”, “*blame others*”, “*self-praise*” and “*praise others*”.

We started to explore the use of the Path Model of Blame [145] for the detection of blame/praise from text. As we are not concerned with the identification of the degree of blame but the existence of blame, there is no need to determine “*Reasons*” as in the original Path Model of Blame. Also, instead of identifying “*Intentionality*”, “*Capacity*” and “*Obligation*”, we replace them with “*Foreseeability*” and “*Coercion*”. According to the path model in Figure 2.6, “*Capacity*” deals with the ability of the moral agent to have known about the actions and its effects beforehand, in other words, its foreseeability. Foreseeability refers to an agent’s foreknowledge about actions and their effects. Clearly, we can say that intentionality entails foreknowledge [151]. Various other papers in this area [151, 152, 230] state that there is a close interplay between intentionality and foreseeability. As such, we replace “*Intentionality*” and “*Capacity*” with “*Foreseeability*”. According to the Webster dictionary to *coerce* is “to make (someone) do something by using force or threats”. In Figure 2.6, obligation deals with the extent to which the moral agent had the ability to prevent the negative event. In this case, the perceiver is considering “could the agent have been forced to carry out the action” or “was the agent tricked into carrying out the actions”. Thus, coercion covers not only cases

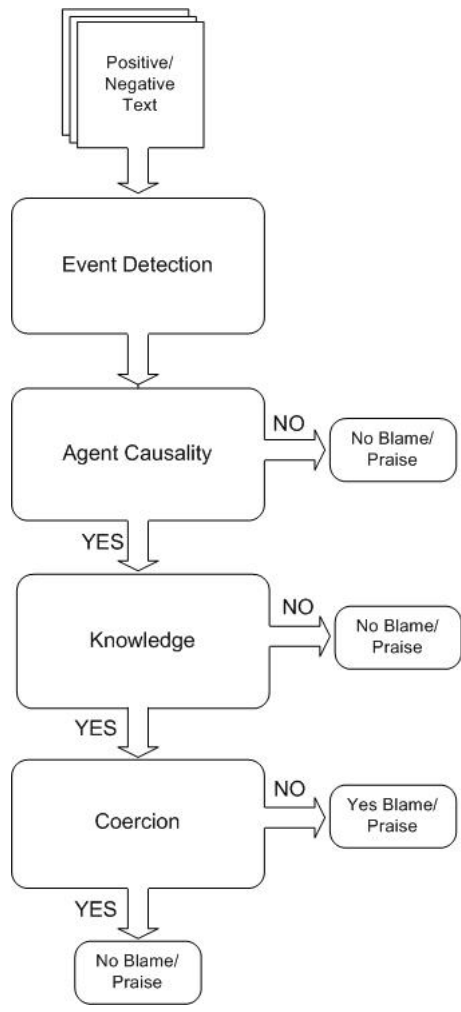


Figure 4.1: Revised Path Model of Blame.

where the agent was forced but also where the agent was tricked to execute the action. Typically coercion is thought to carry with it the implication to diminish the targeted agent’s freedom and responsibility [11]. Thus, we replace “*Obligation*” with “*Coercion*”. The revised Path Model of Blame is illustrated in Figure 4.1.

It is generally acknowledged that a blameworthy entity must be capable of reasoning, and capable of taking a decision. On the off chance that an entity does not have these requirements, it is exempted from blame [267]. Thus we use named entity recognition focusing on entities of persons, organization and country. We also identify the use of pronouns representing persons based on the Part-of-Speech (POS) tagging results.

We first pre-process text by carrying out sentence splitting and tokenisation<sup>2</sup>, POS tagging<sup>3</sup>, word sense disambiguation (WSD)<sup>4</sup>, named entity recognition (NER)<sup>5</sup>, dependency parsing<sup>6</sup>, and polarity detection using majority vote based on the lexicon matching results obtained with three sentiment lexicons as explained in section 3.2.2.3. Negation handling as explained in section 3.2.2.2 is also considered during polarity detection process.

In order to use the revised Path Model of Blame for the detection of blame/praise from text, we need to first detect events and then determine “*Agent Causality*”, “*Foreseeability*” and “*Coercion*”. In the following, we describe how each of the steps can be performed.

#### 4.2.1 Event Detection

We look at the “verb+object” combination as identified using the Stanford dependency parser and take note of the agent of the verb. We use the majority voting mechanism mentioned above for polarity detection. Negatively or positively valenced events are extracted from sentences expressing negative or positive polarity respectively.

In the example shown in Figure 4.2, we see that the event detected by the “verb+object” pattern is “passed exam”. And the agent of the verb “passed” here is “I”. We then detect the polarity of the event by searching for positive or negative words modifying the event taking into account of negation. In this example, the verb “passed” carries a positive polarity. As such, the

---

<sup>2</sup><http://nlp.Stanford.edu/software/tokenizer.shtml>

<sup>3</sup><http://nlp.stanford.edu/software/tagger.shtml>

<sup>4</sup><http://www.nltk.org/>

<sup>5</sup><http://nlp.stanford.edu/software/CRF-NER.shtml>

<sup>6</sup><http://nlp.stanford.edu/software/stanford-dependencies.shtml>



## Stanford Parser Output

*I passed an exam that I was absolutely certain that I had failed.*

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBD passed)
      (NP (DT an) (NN exam))
      (SBAR (IN that)
        (S
          (NP (PRP I))
          (VP (VBD was)
            (ADJP (RB absolutely) (JJ certain))
            (SBAR (IN that)
              (S
                (NP (PRP I))
                (VP (VBD had)
                  (VP (VBN failed))))))))))
      nsubj(passed-2, I-1)
      root(ROOT-0, passed-2)
      det(exam-4, an-3)
      dobj(passed-2, exam-4)
      mark(certain-9, that-5)
      nsubj(certain-9, I-6)
      cop(certain-9, was-7)
      advmod(certain-9, absolutely-8)
      ccomp(passed-2, certain-9)
      mark(failed-13, that-10)
      nsubj(failed-13, I-11)
      aux(failed-13, had-12)
      ccomp(certain-9, failed-13)
```

Figure 4.2: An example dependency parse result.

event is considered as a positively valenced event.

### 4.2.2 Agent Causality

Here, one must establish that a moral agent caused an event. We first make use of a popular explicit intra-sentential pattern for causation expression which is “NP *verb* NP” where NP is a noun phrase [83] and then we identify the agent within the noun phrase. If the intra-sentential pattern is not found we consider verbs in the text that belong to the CAUSE class and the CAUSE-TO semantic relation which are defined in the WordNet. In order for “Agent Causality” taking the value “True”, the agent must be a person entity (including pronouns).

In the example shown in Figure 4.2, we see that the intra-sentential pattern “NP *verb* NP” is present and the dependency parse result shows that verb “passed” is associated with the subject “I” (first person pronoun). This tells us that the agent is a moral agent within the context of the sentence.

For all the self categories (“*self-blame*” or “*self-praise*”), the agent must be a first person pronoun. For other categories (“*blame others*” or “*praise others*”), the agent must not be a first person pronoun, but must be one of the following: a pronoun, a person, country or organization as identified using the NER tool.

### 4.2.3 Foreseeability

This is used to determine intentionality and foreseeability, as both work hand in hand and intentionality entails foreknowledge [152, 152]. It is generally the commitment and foreknowledge to work toward a certain outcome. We rely on a set of verbs which indicate foreseeability. These include verbs of communication as suggested in Mao et al [151] and other verb classes which include verbs of creation, verbs of consumption, verbs of competition, verbs of possession and verbs of motion. These classes of verbs are defined in the WordNet lexnames<sup>7</sup> and can be identified by looking at the WordNet sensekey of the verbs.

The first part of the lexname field is a two digit file number. Followed by the lexicographer filename which the number represents, and then the syntactic category represented by a number (1=NOUN,2=VERB,3=ADJECTIVE,4=ADVERB).

Example: *When I did not **speak** the truth.*

In the example above, the communication verb “speak” indicates that the subject “I” had foreknowledge of the event of “speaking the truth”. Using WSD we can identify the verb speak in the sentence has a sensekey of speak%2:32:03:: and the number 32 indicates that it is a communication verb which includes verbs of asking,telling,ordering,singing.

Figure 4.3 shows a list of the verb lexicographer files we focus on for determining foreknowledge and by extension intentionality. We believe that from the description provided the actions they represent requires the agent to have foreknowledge of said actions.

---

<sup>7</sup><https://wordnet.princeton.edu/man/lexnames.5WN.html>

File Number	Name	Contents
29	verb.body	verbs of grooming, dressing and bodily care
31	verb.cognition	verbs of thinking, judging, analyzing, doubting
32	verb.communication	verbs of telling, asking, ordering, singing
33	verb.competition	verbs of fighting, athletic activities
34	verb.consumption	verbs of eating and drinking
35	verb.contact	verbs of touching, hitting, tying, digging
36	verb.creation	verbs of sewing, baking, painting, performing
38	verb.motion	verbs of walking, flying, swimming
40	verb.possession	verbs of buying, selling, owning
41	verb.social	verbs of political and social activities and events

Figure 4.3: List of lexicographer files with file numbers and a brief description of file's contents.

---

**Algorithm 4:** Sentence level Foreknowledge detection.

---

```

input : Sentence

output: True or False

Tokens = Tokenize(Sentence);
Tags = POSTagger(Sentence);
WordSenseList = empty;
hasForeknowledge= false;

for  $i = 0; i < Tokens.Count; i = i + +$  do
    Token = Tokens[i];
    Tag = Tags[i];
    WordSense = GetWordSenseOfWordinSentence(Sentence,Token,Tag);
    WordSenseList.Add(WordSense);
end

for  $i = 0; i < WordSenseList.Count; i = i + +$  do
    if VerbIsInForseeabilityClass(WordSenseList[i]) then
        | hasForeknowledge = true;
    end
end

return hasForeknowledge;

```

---

#### 4.2.4 Coercion

To identify coercion, we look at the extension verb classes presented in [118] focusing on verbs in the URGE (13 members), FORCE (46 members) and FORBID (17 members) classes.

Example: *I was **forced** to quit the job in the city.*

In the example above, using word sense disambiguation, the verb “forced” is of sense “*to cause to do through pressure or necessity, by physical, moral or intellectual means*”. The agent “I” in this case did not willingly quit the job and the sentence does not mention who forced the agent. Thus, the sentence is classified as “*Others*” (i.e., no blame or praise).

example

```
{  
  "word": "forced",  
  "name": "coerce.v.01",  
  "key": "coerce%2:41:00::",  
  "pos": "VBN",  
  "index": 2  
}
```

---

**Algorithm 5:** Sentence level Coercion detection.

---

```
input : Sentence

output: True or False

Tokens = Tokenize(Sentence);
Tags = POSTagger(Sentence);
WordSenseList = empty;
hasCoercion= false;

for  $i = 0; i < Tokens.Count; i = i + +$  do
    Token = Tokens[i];
    Tag = Tags[i];
    WordSense = GetWordSenseOfWordinSentence(Sentence,Token,Tag);
    WordSenseList.Add(WordSense);
end

for  $i = 0; i < WordSenseList.Count; i = i + +$  do
    if VerbIsInCoercion(WordSenseList[i]) then
        | hasCoercion = true;
    end
end

return hasCoercion;
```

---

### 4.3 Corpus Creation

We created our data from the ISEAR dataset, which was collected during the 1990s by a large group of psychologists by asking nearly 3,000 participants from different cultural background about their emotional experiences. This dataset contains 7,660 comments, each of which is labelled with one of the seven emotions (*joy, fear, anger, sadness, disgust, shame* and *guilt*).

We asked two English-speaking individuals to annotate each comment in the ISEAR dataset as “*blame*”, “*praise*” or “*others*”. For comments expressing *blame* or *praise*, the annotators further labelled them as “*self-blame*”, “*blame others*”, and “*self-praise*” and “*praise others*”. The annotators were provided with the annotation guidelines and sample annotation results. A web-based interface has been developed to ease the task of annotation. We did not provide them with the ISEAR emotion

labels of the comment as we believe that the emotion label information, although might be helpful, will create bias and influence the annotators.

**Sentences:**  
When I realized that I was directing the feelings of discontent with myself at my partner and this way was trying to put the blame á on him instead of sorting out my own feelings.

Sentence Type  
Blame ▾

Sentence Class  
Self Blame (blaming one's self for something) ▾

**Submit**

Figure 4.4: Annotation User Interface.

The inter-annotator agreement for our data set is shown in Table 4.1. There are several agreement measures which have been proposed in the literature [13], We measure the reliability of the annotation results by using the kappa ( $k$ ) coefficient [121], which is defined as  $k = A_o - A_e / 1 - A_e$  where  $A_o$  is the observed agreement, and  $A_e$  is the expected agreement by chance. We obtained a  $k$  score of 0.62. Using the scales for interpreting Kappa provided in [128] and [85] in terms of strength of agreement, our score can be interpreted as a good and substantial agreement.

Annotation guidelines and instructions provided to our annotators is available in Appendix B.

Table 4.1: Annotation agreement matrix.

		Annotator 1			Total
		Blame	Praise	Others	
Annotator 2	Blame	<b>3483</b>	222	279	3984
	Praise	227	<b>778</b>	305	1310
	Others	348	299	<b>1719</b>	2366
	Total	4058	1299	2303	7660

It can be observed from Table 4.1 that, 45% of the comments in ISEAR are labeled as “*blame*” and 10% as “*praise*”. We only keep the comments where both annotators reach an agreement. On fine grained labelling, we had a discrepancy of about 17%. We then got both annotators to re-examine these 17% comments to reach an agreement. Our final dataset consists of 57.1% self blames and 42.9% blames directed towards others within the blame context; and 66.2% self praises and 33.8% praises directed towards others in the praise context.

#### 4.4 Experiments and Analysis

Our approach uses a sentence level detection. A comment can contain 1 or more sentences, each comment is split into individual sentences, we then classify each sentence to blame, neutral or praise. The categories are then weighed and the category with the highest weight is allocated to the comment. Neutral sentences carry no weight.

$$Cw = \frac{Sc}{St}$$

$Cw$  = Category Weight

$Sc$  = Number of Sentences with class

$St$  = Total Number of Sentences

In the event that we have equal number of praise and blame sentences, we categorize the comment as blame as negative emotions outweigh positive ones. In this section, we present the evaluation results of our blame detection approach and compare it with supervised learning approaches trained on the “bag-of-words” features. Experiments for the supervised classifiers were carried out using Weka<sup>8</sup> with documents pre-processed with stopword removal. We report the results using 10-fold cross validation. Note that such a comparison is not fair since our approach does not make use of any labelled data.

Class	NB			SVM			Our Approach		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
<i>blame</i>	<b>0.74</b>	0.60	0.66	0.73	<b>0.80</b>	<b>0.76</b>	0.73	0.76	0.75
<i>praise</i>	0.40	0.38	0.39	<b>0.60</b>	0.44	0.51	0.43	<b>0.63</b>	<b>0.52</b>
<i>others</i>	0.47	<b>0.66</b>	0.55	0.61	0.57	<b>0.59</b>	<b>0.63</b>	0.53	0.58

Table 4.2: Classification results of *blame*, *praise* or *others*.

It can be observed from Table 4.2 that supervised SVM performed the best in classifying *blame*, but only slightly outperforms our approach by about 1% in F-measure. Supervised NB gives much worse results with F-measure lower than that of SVM. Our approach achieves similar performance as SVM on the *praise* category and outperforms NB by 13% in F-measure. The dataset has a higher

<sup>8</sup><http://www.cs.waikato.ac.nz/ml/weka/>

number of negative comments and this is reflected in the results obtained on the *blame* category having better performance than those from *praise* across all classifiers.

For fine-grained classification, it can be observed from Table 4.3 that SVM performed better than NB on all categories. However, our approach performed better than both SVM and NB. In classifying into *self-blame* (blame directed towards oneself) and *blame others* (blame directed towards other people), our approach performs better than SVM and NB with an average F-measure difference of about 6% and 12% respectively. In the *self-praise* and *praise others* categories, our approach performs better than both SVM and NB with an average F-measure difference of about 20% and 17%.

Class	NB			SVM			Our Approach		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
<i>self-blame</i>	0.54	0.40	0.46	0.52	0.58	0.55	<b>0.58</b>	<b>0.59</b>	<b>0.59</b>
<i>blame others</i>	0.43	0.49	0.46	0.50	0.47	0.48	<b>0.53</b>	<b>0.68</b>	<b>0.57</b>
<i>self-praise</i>	0.44	0.42	0.43	<b>0.54</b>	0.38	0.45	0.49	<b>0.51</b>	<b>0.50</b>
<i>praise others</i>	0.16	0.33	0.25	0.22	0.12	0.16	<b>0.49</b>	<b>0.53</b>	<b>0.51</b>
<i>others</i>	0.51	0.52	0.51	0.56	<b>0.62</b>	<b>0.59</b>	<b>0.63</b>	0.53	0.58

Table 4.3: Classification results of *self-blame*, *self-praise*, *blame-others*, *praise-others* or *others*.

#### 4.4.1 Error Analysis

Error analysis was conducted on the blame detection results in order to better understand how the proposed approach performs. Table 4.4 shows some sample detection results. Within each step in our approach, we log instances where an error may have occurred or a failure to proceed to the next step is detected or if WSD returns no results or polarity detection is neutral or misclassification. In general, whenever the output of one process is the input of another process we track it and log instances where a failure or error may have occurred. We select random error samples of about 5 records for each stage where an error has occurred and examined them to identify causes of errors and possible mitigation. In total, we reviewed a total of 20 error instances in the dataset.



Table 4.4: Detection examples of correct and failed detection using our approach on the dataset

<b>Example</b>	<b>Expected Outcome</b>	<b>Outcome</b>
1. <i>When I learned that several people had died in the street due to the cold weather.</i>	Blame	Neutral
2. <i>I had to decline an appointment which had been very important for the other person. For this reason we even got into trouble.</i>	Blame	Blame
3. <i>When I am in an environment or with a person much worse off than me, I realize how privileged I am.</i>	Blame	Neutral
4. <i>I happened to overhear something which I was not meant to hear.</i>	Blame	Neutral
5. <i>When I got home from a pleasant trip abroad, I got to know that I had been accepted at university.</i>	Pride	Pride
6. <i>My elder sister forced me to do a few things which I did not like to do.</i>	Blame	Neutral

In example 1, we failed to identify blame as the model requires that the agent to be blamed must meet moral requirement. However, the agent responsible for the event "death of people" in this case is the "weather" and that cannot be responsible for blame. Thus, there is a huge distinction in how the model attributes blame when compared to how the average person attributes blame. This type of error accounted for about 32% of the overall errors encountered. Our model would benefit greatly if the model was more inline with how human beings attribute blame to natural events (non-moral agents).

In example 3, we failed to identify blame as the model requires that an event to be detected in the sentence. Our approach failed to identify an event using the VERB + OBJECT relationship and as a result failed to detect blame/praise. This type of error accounted for about 27% of the overall errors encountered. We have begun investigating how to handle sentences which do not have an outright OBJECT. Perhaps looking into the handling of transitive and intransitive verbs, we can also look at analysing noun phrases in the absence of OBJECTS.

In example 4, we failed to identify blame as the model requires that a non-neutral event must be

detected. Our system failed to identify any positive or negative polarity in the sentence, as such the sentence was classified as Neutral. Inability to detect the polarity of a sentence accounted for about 13% of the overall errors encountered. To mitigate the chances of a polar term not been correctly identified, we decided to use a majority vote approach with 5 lexical resources. Perhaps we could consider a hybrid approach for polarity detection.

In example 6, we failed to identify blame as the model requires that for blame to exist the agent must be blameworthy. Our system did not classify the sentence as blame as it contained coercion of the agent and as such the sentence is classified as Neutral. We don't classify these as errors because they are cases where the model as identified the presence of coercion and as such the agent is exempt from blame.

ISEAR dataset contains personal experience expressed by a wide range of participants and hence might contain lots of informal and ill-grammatical text, we can say that our approach performs reasonably well with such dataset. Our approach relies on results generated from a series of NLP tasks such as POS tagging, word-sense disambiguation, dependency parsing and polarity detection in order to be able to assign values to a set of variables for blame/praise detection. Thus, any error that occurs will be propagated down the pipeline process. Furthermore, failure in detecting the polarity of text will make it impossible for our approach to identify the underlying blame/praise category. However, our approach performs reasonably well especially on the fine-grained classification of detecting the directions of blames or praises. This is very useful in not only identifying the entity responsible but also for inferring emotions such as guilt, remorse, anger and many more.

For example, for the sentence *“When I caused problems for somebody because he could not keep the appointed time and this led to various consequences.”*, the Agent (“I”) is blameworthy and hence we can infer that the sentence expresses an emotion of *guilt*.

In this chapter, we presented our work on blame/praise detection in text based on the “Path Model of Blame”, with the aim of addressing our second research question: *Can we detect praise and blame in text, elements necessary for the detection of emotions such as guilt, remorse and many more?*.

In order to answer this question, we created and annotated a dataset for blame/praise detection. We proposed an approach adapted from the Path Model of Blame and used this adaptation for detecting expressions of blame and praise in text. Experimental results on our dataset show that

our approach gives similar performance compared to supervised classifiers when classifying text as *blame* or *praise*. For fine-grained classification of identifying the direction of blame and praise, our approach outperforms the supervised methods by a large margin of 14% in F-measure compared to NB and 13% in F-measure compared to SVM.

## Chapter 5

# Content-Based Conflict of Interest Detection for Articles in Wikipedia

In this chapter, we discuss an application area for implicit emotion detection, we believe that implicit emotion detection may have potential usefulness for conflict-of-interest (CoI) detection in Wikipedia articles. We build a CoI corpus and explore various types of features including linguistic and stylometric features, presentation features, bias features and emotion features for CoI detection. We also observe from the collected data, that 55% of our dataset were articles on company profiles and 43% articles on individual profiles and 2% others. With only 21% anonymous edits

### 5.1 Introduction

Reference works writers put in a lot of effort to keep the language as unbiased as possible, in other to maintain the objective perspective of such platforms. This has not always been the case as you find that the writers and contributors often use other ways to express affect without using explicit opinionated vocabulary. For example, Wikipedia's policy called neutral point of view (NPOV), which states that "articles should represent fairly, proportionately, and as far as possible without bias, all significant views that have been published by reliable sources". It is then crucial to understand the nature of bias, its linguistic realization and perhaps automatically detect it.

Take two example documents from our Wikipedia dataset, one is an article classified as CoI while the other is not:

- **CoI example:** *Kaizaad Kotwa, born in Mumbai, India, is an award winning professor and writer, actor, director, producer and designer. Currently he is a professor of theatre and film at Ohio State University in Columbus, Ohio. He recently won the Griffin Society Award for*

*Best Professor and in 2007 was named one of the top professors in Ohio. He is the co-owner and co-Artistic Director of Poor-Box Productions, along with his mother Mahabanoo Mody-Kotwal, a famous actor, director and producer in India.*

- **Non-CoI example:** *“Enrica Zunic” is the pseudonym of “Enrica Lozito”, an Italian science-fiction writer. She lives and works in Turin. Her work is partly inspired by her activities with Amnesty International. In 2003 she won the Premio Italia award for science fiction.*

Using our proposed approach, a number of interesting features are identified as shown in Figure 5.1. It can be observed that the CoI example when compared to the non-CoI one contains more subjective sentences, bias sentences, emotion and more praise/blame expressions.

Our main aim in this work is to detect CoI articles based solely on the content of the articles without relying on any related metadata. We explore a rich set of features including stylometric features, the presentational features by focusing on the existence of Rhetorical Structure Theory’s (RST’s) presentational relations, various forms of language biases and implicit/explicit emotions. We then investigate using different combinations of features to train supervised binary classifiers for CoI detection. Our results show that the best result of 0.67 in F-measure is obtained when training Support Vector Machines (SVMs) from a combination of all features. Also, further combining various features with document-level representations either in the form of bag-of-words or dense representations by combining pre-trained word vectors does not bring any performance gains. As we only have the labeled CoI class, but not the non-CoI class, we have also explored the use of one-class classification for CoI detection. The results show that using stylometric features outperforms other types of features or a combination of them. Also, one-class classifier gives higher precision values compared to binary classifiers. To the best of our knowledge, we are the first to carry out automatic CoI detection on Wikipedia articles based solely on text content.

Our main contributions are summarised below:

- We have built a CoI dataset which contains 3,280 CoI articles and 3,450 non-CoI articles, which could be used in future research on CoI detection;
- We have proposed a set of features based on our research of existing work close to CoI detection and analysis of the data collected and have identified the most effective features through

Kaizaad Kotwal, born in Mumbai, India, is an award winning professor and writer, actor, director, producer and designer.

**Emotion:** joy, trust, anticipation, surprise  
**Sentiment:** positive  
**Bias:** 0.056  
**Praise/Blame:** Praise  
**Type:** Active Sentence

Currently he is a professor of theatre and film at Ohio State University in Columbus, Ohio.

**Emotion:** trust  
**Sentiment:** Positive  
**Bias:** 0.018  
**Praise/Blame:** Neutral  
**Type:** Active Sentence

He recently won the Griffin Society Award for Best Professor and in 2007 was named one of the top professors in Ohio.

**Emotion:** joy, trust, anticipation, surprise  
**Sentiment:** Positive  
**Bias:** 0.19  
**Praise/Blame:** Praise  
**Type:** Active/Passive Sentence

He is the co-owner and co-Artistic Director of Poor-Box Productions, along with his mother Mahabanoo Mody-Kotwal, a famous actor, director and producer in India.

**Emotion:** trust, joy  
**Sentiment:** Positive  
**Bias:** 0.083  
**Praise/Blame:** Praise  
**Type:** Active Sentence

(a) CoI example article

“Enrica Zunic” is the pseudonym of “Enrica Lozito”, an Italian science-fiction writer.

**Emotion:** Neutral  
**Sentiment:** Neutral  
**Bias:** 0  
**Praise/Blame:** Neutral  
**Type:** Active Sentence

She lives and works in Turin.

**Emotion:** Neutral  
**Sentiment:** Neutral  
**Bias:** 0  
**Praise/Blame:** Neutral  
**Type:** Active Sentence

Her work is partly inspired by her activities with Amnesty International.

**Emotion:** joy  
**Sentiment:** Positive  
**Bias:** 0.181  
**Praise/Blame:** Neutral  
**Type:** Passive Sentence

In 2003 she won the Premio Italia award for science fiction.

**Emotion:** trust, surprise, joy, anticipation  
**Sentiment:** Positive  
**Bias:** 0.1  
**Praise/Blame:** Praise  
**Type:** Active Sentence

(b) Non-CoI example article

Figure 5.1: Two sample documents with features identified by our approach. Words/phrases underlined in text are those which be found in an emotion or sentiment lexicon. Due to space constraint, we only show some key features here such as the emotion/sentiment label, bias score, praise/blame indicator, and sentence type.

extensive experiments on our CoI dataset;

- We have also investigated the effectiveness of using one-class classification for CoI detection.

## 5.2 Our Approach

We address the CoI detection problem as binary classification which determines if a given document belongs to the category of CoI or non-CoI. We make the following hypotheses:

1. Since CoI is a sub-category of the “NPOV disputes” Wikipedia category, CoI articles inherit various linguistic and stylometric characteristics from their parent Wikipedia categories including those typically found in vandalism and bias;
2. CoI articles contain more subjective sentences than non-CoI articles;
3. The presentation of content in CoI articles will tend to increase the reader’s interest/regard for the subject matter;
4. Since the choice of words projects opinions and preferences, CoI articles likely contain more expressions of implicit or explicit emotions.

In this section, we explore a rich set of features to test our hypotheses above and to train supervised classifiers for CoI detection. Figure 5.2 shows the work flow of our approach.

### 5.2.1 Stylometric Features

Stylometric features these try to recognise patterns or writing styles in text. This technique has been applied in the area of authorship attribution [216, 245, 12], opinion mining [195], and forensic linguistics [270, 189]. We create a list of features selected from previous research work in vandalism and bias as mentioned in the Related Work section. Since not all features are relevant to our CoI detection task, We perform feature selection using the implementation of InfoGain and Chi-Square available in Weka<sup>1</sup> to eliminate insignificant features. We also include the nine universal dependency groups<sup>2</sup>, detection of which is done using the Stanford Dependency Parser<sup>3</sup>. The final set of features is listed in Table 5.1. This set of features is relating to Hypothesis 1.

---

<sup>1</sup><http://www.cs.waikato.ac.nz/ml/weka/>

<sup>2</sup><http://universaldependencies.org/docsv1/u/dep/index.html>

<sup>3</sup><http://nlp.stanford.edu/software/stanford-dependencies.shtml>

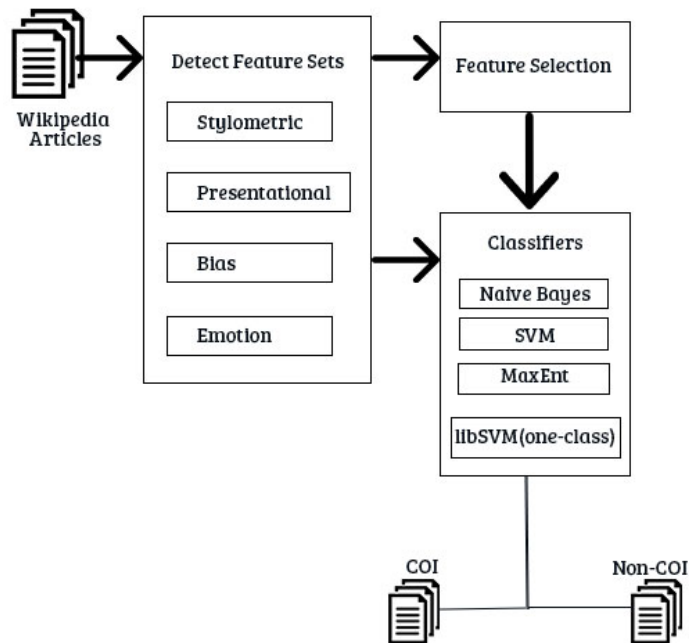


Figure 5.2: The overall work flow of our proposed approach.

### 5.2.2 Bias Features

In Recasens et al [214], two major classes of bias in Wikipedia edits have been discussed, *framing bias* and *epistemological bias*. The former is related to subjective words and phrases that state a particular point of view, while the latter deals with linguistic features which are related to the believability of a proposition. We use the same classes of bias as discussed in their work [214] and identify the existence of the classes in a Wikipedia article based on a bias lexicon<sup>4</sup>. We also consider other words/phrases which may introduce bias as illustrated in the Wikipedia’s manual of style/Words to Watch<sup>5</sup>. The bias features are shown in Table 5.2. This set of features is relating to Hypothesis 1 and 2.

### 5.2.3 Presentational Features

Rhetorical Structure Theory (RST) is a discourse theory, which offers an explanation of the coherence of texts. It provides a way to describe the relations among text and has been used to successfully analyse a variety of text types [254, 255]. In RST, presentational relations are relations that tend to increase the inclination in the reader or to increase the acceptance of the content [148] by the reader.

<sup>4</sup>[http://www.mpi-sws.org/~cristian/Biased\\_language.html](http://www.mpi-sws.org/~cristian/Biased_language.html)

<sup>5</sup>[https://en.wikipedia.org/wiki/Wikipedia:Manual\\_of\\_Style/Words\\_to\\_watch](https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Words_to_watch)



Table 5.1: Stylometric features.

Feature Name	Description
<i>Sentence Level</i>	
Average_Sentence_Length	Average length of the sentences in the document
Average_Unique_Word_Count	Average # of unique words per sentence
Average_Punctuation	Average number of punctuations per sentence
Adjective_Rate	Rate of adjectives per sentence
CC_Rate	Rate of coordinating conjunctions per sentence
Pronouns_Rate	Rate of pronouns per sentence
Word_Count_Score	Total # of words / Total # of sentences
Unique_POS_per_Sentence	Rate of unique Part-of-Speech (POS) tags per sentence
<i>Document Level</i>	
Sentence_Count	Total # of sentences in the document
Unique_Word_Count	Total # of unique words in the document
No_of_Verbs	Total # of verbs in the document
No_of_CC	Total # of coordinating conjunctions in the document
No_of_CompAdverbs	Total # of comparative adverbs in the document
No_of_Adjectives	Total # of adjectives in the document
Special_clausal_dependents	Total # of special clausal dependents in the document
Active_Sentences	Total # of non-passive sentences
Non_core_dependents_of_clausal_predicates	Total # of non-core dependents of clausal predicates
Core_dependents_of_clausal_predicates	Total # of core dependents of clausal predicates
Noun_dependents	Total # of Noun dependents
Compounding_and_unanalyzed	Total # of Compounding and unanalyzed dependencies
Case-marking, prepositions, possessive	Total # of Case-marking, prepositions, possessive
Coordination	Total # of Coordination dependencies
Loose_joining_relations	Total # of loose joining relations
Sentence_head_and_Unspecified_dependency	Total # of Sentence head and Unspecified dependency
Complexity_Score	Text complexity score

Subject matter relations are relations whose intended effect is that the reader recognises the relation in question [149].

We focus our work on identifying the existence of presentational relations<sup>6</sup> using cue words as relation signals. We use 10 presentational relations as shown in Table 5.3, as they increase readers' acceptance of text in one form or the other. We built a simple cue phrase detector with phrases provided in various RST research [254] and relation nucleus/satellite positioning described in [148]. This set of features is relating to Hypothesis 3.

#### 5.2.4 Emotion Features

We focus on Ekman's six basic emotions (*joy, sadness, anger, surprise, fear, disgust*) and implement both explicit and implicit emotions detection. Emotions can be expressed explicitly by using

<sup>6</sup><http://www.sfu.ca/rst/01intro/definitions.html>

Table 5.2: Bias features and subtypes.

Bias	Subtypes
Epistemological	Factive verbs / Entailments / Assertives / Hedges
Framing	Subjective terms / Intensifiers
Others	Puffery / Contentious labels / Unsupported attributions / Expressions of doubt / Editorialising

Table 5.3: Definitions of 10 presentational relations used (N stands for nucleus, R for reader and W for writer).

Relation Name	Intention of W
Antithesis	R’s positive regard for N is increased
Background	R’s ability to comprehend N increases
Concession	R’s positive regard for N is increased
Enablement	R’s potential ability to perform the action in N increases
Evidence	R’s belief of N is increased
Justify	R’s readiness to accept W’s right to present N is increased
Motivation	R’s desire to perform action in N is increased
Preparation	R is more ready, interested or oriented for reading N

“emotion-bearing words” or implicitly without such words. For explicit emotions, we use a simple lexicon-based approach with negation handling based on a modified version of the NRC lexicon [173]; and for implicit emotions, we use the rule-based approach [272]. In addition, we also perform polarity detection (*positive* and *negative*) using majority voting based on the lexicon matching results obtained with three sentiment lexicons, SentiWordNet [70], AFINN [95] and the Subjectivity Lexicon [287]. We implement a contextual valence shifter as described in [206] to detect polarity change in context. Apart from emotion and polarity features, we also consider the expressions of *blame* and *praise* as additional features using the method proposed in [190] for detection. This set of features is relating to Hypothesis 4.

## 5.3 Experiments

### 5.3.1 Data

We construct our dataset by collecting 4,050 articles from Wikipedia which have been categorised as conflict of interest (CoI) items<sup>7</sup>. This CoI category is a sub-category of “NPOV disputes”. Wikipedia encourages its editors to pick an article from this category and decide whether it meets its notability policy<sup>8</sup>. If one believes the article should be kept, he/she needs to review the text to

<sup>7</sup>[https://en.wikipedia.org/wiki/Category:Wikipedia\\_articles\\_with\\_possible\\_conflicts\\_of\\_interest](https://en.wikipedia.org/wiki/Category:Wikipedia_articles_with_possible_conflicts_of_interest)

<sup>8</sup><https://en.wikipedia.org/wiki/Wikipedia:Notability>

ensure that it complies with NPOV. This human categorisation of Wikipedia articles will be our basis for evaluating our results.

In order to build a dataset containing both CoI and non-CoI articles, for each CoI article, we randomly select non-CoI articles from its first associated Wikipedia category. For example, a CoI article might be associated with two categories, “1932 births” and “Living people”. We randomly select a non-CoI article from the category “1932 births”. This resulted in a total of 4,600 non-CoI articles selected from over 100 Wikipedia categories. We have considered various criteria for the selection of non-CoI articles such as the age of article, number of views, editor information. We found that identifying a threshold on these meta-data properties which cuts across the various Wikipedia categories and Wikipedia sectors would require a fine-tooth comb. For example, an article maybe older but has fewer views than a newer article OR articles from a particular category may have more views than other categories. As a result, we chose the random selection approach as long as the article was from the same category as a CoI article and did not belong to CoI disputes category. We focus on the article content as our means of classification and ignore the meta information provided by Wikipedia such as the editor(s) of a Wikipedia edit, time and date of creation, associated IP address, etc. Our dataset is made available at<sup>9</sup>

### 5.3.2 Preprocessing

We pre-process the dataset by removing the top 1% articles and the lower 5% of the articles based on the document length. This reduces the total number of documents to 3,280 CoI articles and 3,450 non-CoI articles. The vocabulary size for the dataset is 52,302. We then carry out sentence splitting and tokenisation, stopword removal, stemming and remove words occurred less than ten times. For implicit emotion detection and blame/praise detection, we also perform part-of-speech (POS) tagging using the Stanford POS Tagger<sup>10</sup>, word sense disambiguation (WSD) using the classic Lesk algorithm for WSD in NLTK<sup>11</sup>, and dependency parsing using the Stanford Dependency Parser.

To represent documents, apart from the commonly used bag-of-words approach, we also consider using doc2vec [130] which modifies the word2vec algorithm [160] for unsupervised learning of continuous representations for larger blocks of text, such as sentences, paragraphs or entire docu-

---

<sup>9</sup>The dataset will be made available upon publication.

<sup>10</sup><http://nlp.stanford.edu/software/tagger.shtml>

<sup>11</sup><http://www.nltk.org/howto/wsd.html>

ments. Recent work in the area of NLP has shown it to be a strong alternative for both bag-of-words and bag-of-n-grams models. We use Gensim<sup>12</sup> which has an implementation of doc2vec. We ignore words occurred less than 10 times and generate a vector representation of each article using the pre-trained vectors from the Google News dataset<sup>13</sup> with about 100 billion words, 300-dimensional vectors. The size of the context window we use is 3 before and after the predicted word. The final generated document vectors have 100 dimensions.

### 5.3.3 Feature Selection

Here we aim to identify the features that are mostly useful for prediction of CoI. The guiding idea is that a good feature set should contain features that are highly correlated with the class, yet uncorrelated to each other. We use Correlation-based Feature Subset Selection (CFS) and Information Gain Ratio (IGR) to rank features on all our feature sets from the training set. Table 5.4 shows the top ranked features by CFS and IGR respectively. It is interesting to see that the top 3 features (*Blame*, *Praise* and *Polarity\_Score*) returned by CFS all belong to the *Emotion* category. These three features are ranked among the top 12 positions by IGR. This indicates that CoI articles tend to contain more expressions of “Blame” or “Praise” and show clearer polarity compared to non-CoI articles. We also see that *Active\_Sentences*, being ranked at 4th by CFS and 3rd by IGR, is an important discriminative feature for CoI detection. *Average\_Sentence\_Length* and *Average\_Unique\_Word\_Count*, ranked at the 7th and 8th positions by both CFS and IGR, are another two stylometric features that are important for CoI classification. Other common stylometric features ranked among the top 15 positions by both CFS and IGR include *Non\_core\_dependents\_of\_clausal\_predicates*, *No\_of\_CC*, *Adjective\_Rate*, *CC\_Rate* and *Sentence\_Count*. This shows that CoI articles tend to use more coordinating conjunctions and adjectives, and have more sentences compared to non-CoI articles.

In summary, among the top 15 features ranked by IGR and CFS, 11 are the same (73%). The merged features obtained by each individual method are listed in Table 5.5. Most of the top features are Stylometric features (74%) followed by the Emotion (21%) and Bias (5%) features. We also found that no Presentational features appear in the top 15 positions. The feature selection results indicate that stylometric features are very important in determining whether an article should be classified as CoI. Among various emotion features, *Blame*, *Praise*, *Polarity\_Score* and *Suprise* seem

---

<sup>12</sup><http://radimrehurek.com/gensim/>

<sup>13</sup><https://code.google.com/archive/p/word2vec/>

Table 5.4: Individual feature ranking results from Correlation-based Feature Subset Selection (CFS) and Information Gain Ratio (IGR).

Rank	CFS Ranked Features	IGR Ranked Features
1	Blame	No_of_CC
2	Praise	Sentence_Count
3	Polarity_Score	Active_Sentences
4	Active_Sentences	Non_core_dependents_of_clausal_predicates
5	Non_core_dependents_of_clausal_predicates	Unique_POS_per_Sentence
6	Coordination	Polarity_Score
7	Average_Sentence_Length	Average_Sentence_Length
8	Average_Unique_Word_Count	Average_Unique_Word_Count
9	No_of_CC	Praise
10	Adjective_Rate	CC_Rate
11	CC_Rate	Complexity_Score
12	Pronouns_Rate	Blame
13	Word_Count_Score	Adjective_Rate
14	Sentence_Count	Surprise
15	Bias_Score	Special_clausal_dependents

more important than others. The *Bias\_Score* is also relevant, but less important compared to many Stylometric or some Emotion features. Presentational features do not seem to contribute much to CoI detection. We merge the top 15 ranked features from both CFS and IGS and form our best feature set.

Table 5.5: Merged features from the feature selection results from Correlation-based Feature Subset Selection (CFS) and Information Gain Ratio (IGR).

Feature	Set	Description
Blame	<i>Emotion</i>	Total # of expressions of “Blame”
Praise	<i>Emotion</i>	Total # of expressions of “Praise”
Polarity_Score	<i>Emotion</i>	Aggregated polarity score of the document
Surprise	<i>Emotion</i>	Total # of expressions of “Surprise”
Active_Sentences	<i>Stylometric</i>	Total # of non-passive sentences
Non_core_dependents_of_clausal_predicates	<i>Stylometric</i>	Total # of non-core dependents of clausal predicates
Average_Sentence_Length	<i>Stylometric</i>	Average length of sentences in the document
Average_Unique_Word_Count	<i>Stylometric</i>	Average # of unique words per sentence
No_of_CC	<i>Stylometric</i>	Total # of coordinating conjunctions in the document
CC_Rate	<i>Stylometric</i>	Rate of coordinating conjunctions per sentence
Adjective_Rate	<i>Stylometric</i>	Rate of adjectives per sentence
Pronouns_Rate	<i>Stylometric</i>	Rate of pronouns per sentence
Sentence_Count	<i>Stylometric</i>	Total # of sentences in the document
Coordination	<i>Stylometric</i>	Total # of Coordination dependencies
Word_Count_Score	<i>Stylometric</i>	Total # of words / Total # of sentences
Unique_POS_per_Sentence	<i>Stylometric</i>	Rate of unique Part-of-Speech (POS) tags per sentence
Complexity_Score	<i>Stylometric</i>	Text complexity score
Special_clausal_dependents	<i>Stylometric</i>	Total # of special clausal dependents
Bias_Score	<i>Bias</i>	Aggregated bias score of the document

### 5.3.4 Binary Classification Results

We train supervised classifiers including Support Vector Machines (SVMs), Maximum Entropy (MaxEnt) and Naïve Bayes (NB) using various feature sets and different combinations of them.

10-fold cross-validation is used and the results are averaged over 10 such runs.

We can observe from Table 5.6 that among the four feature sets, *Stylometric* gives the best performance followed by *Emotion*. This is consistent with our feature selection results discussed in Section 5.3.3. It also confirms our hypothesis that the writing styles of editors of CoI articles are similar. *Bias* and *Presentational* features appear to be less useful. This shows that CoI is more than just bias. Presentational features had no member appeared in the top 20 features ranked by CFS or IGR. Although SVM or MaxEnt trained from *Presentational* or *Bias* features give much worse results compared to other feature sets, NB trained from these two types of features sets performs only slightly worse than trained from *Stylometric* or *Emotion* features.

We have also tried combinations of different features sets. For both SVM and MaxEnt, the best performance is given by *All features*. SVM achieves much higher recall than precision with an overall F-measure of 0.67. MaxEnt gives more balanced precision and recall values, but with slightly worse F-measure compared to SVM. We also notice that using *Best features* as listed in Table 5.5 does not lead to improved performance for SVM or MaxEnt. However, the *Best features* set boosts the recall value to 0.94 for NB, although it only gives the precision value of 0.51.

Table 5.6: Conflict-of-Interest (CoI) detection results in Precision, Recall and F-measure using SVM, MaxEnt and NB with various feature sets.

Feature Sets	SVM			MaxEnt			NB		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
Stylometric	0.57	0.74	0.64	0.60	0.61	0.60	0.50	0.91	0.65
Presentational	0.07	0.06	0.06	0.57	0.36	0.44	0.49	0.90	0.63
Bias	0.18	0.16	0.17	0.57	0.26	0.35	0.48	0.91	0.63
Emotion	0.26	0.16	0.24	0.56	0.39	0.46	0.49	0.92	0.64
Stylometric+Emotion	0.57	0.73	0.64	0.60	0.61	0.61	0.50	0.91	0.65
Stylometric+Emotion+Bias	0.57	0.73	0.64	0.61	0.61	0.61	0.50	0.91	0.65
All features	0.58	<b>0.81</b>	<b>0.67</b>	<b>0.64</b>	<b>0.67</b>	<b>0.66</b>	<b>0.62</b>	0.51	0.56
Best features	<b>0.61</b>	0.30	0.40	0.61	0.65	0.63	0.51	<b>0.94</b>	<b>0.66</b>
Bag of Words (BOW)	0.62	0.61	0.61	0.62	0.61	0.62	0.60	0.70	0.64
BOW + Stylometric	0.62	0.62	0.62	0.61	0.62	0.62	0.58	0.76	0.66
BOW + Presentational	0.62	0.62	0.62	0.62	0.61	0.62	0.60	0.71	0.65
BOW + Bias	0.61	0.61	0.61	0.62	0.61	0.62	0.60	0.70	0.65
BOW + Emotion	0.61	0.62	0.61	0.62	0.61	0.61	0.59	0.71	0.65
BOW + Stylometric+Emotion	0.62	0.61	0.62	0.63	0.61	0.62	0.52	0.90	0.66
BOW + Stylometric+Emotion+Bias	0.62	0.62	0.62	0.63	0.61	0.62	0.51	0.90	0.65
BOW + All features	0.62	0.63	0.62	0.62	0.62	0.62	0.57	0.79	0.66
BOW + Best features	0.62	0.62	0.62	0.62	0.61	0.61	0.59	0.73	0.66
doc2vec	0.60	0.64	0.62	0.61	0.62	0.62	0.53	0.67	0.59
doc2vec + Stylometric	0.63	0.63	0.63	0.62	0.63	0.62	0.53	0.71	0.60
doc2vec + Presentational	0.60	0.63	0.61	0.60	0.61	0.61	0.53	0.68	0.59
doc2vec + Bias	0.61	0.63	0.62	0.62	0.62	0.62	0.53	0.69	0.59
doc2vec + Emotion	0.60	0.64	0.62	0.61	0.62	0.61	0.53	0.69	0.60
doc2vec + Stylometric+Emotion	0.63	0.63	0.63	0.62	0.63	0.62	0.53	0.72	0.61
doc2vec + Stylometric+Emotion+Bias	0.63	0.63	0.63	0.62	0.62	0.62	0.53	0.73	0.61
doc2vec + All features	0.62	0.62	0.62	0.61	0.61	0.61	0.53	0.74	0.62
doc2vec + Best features	0.53	0.76	0.63	0.59	0.53	0.55	0.54	0.79	0.64

We have next experimented with document representations using Bag-of-Words (BOW) weighted by TFIDF or doc2vec, and a combination of BOW or doc2vec with various feature sets. The results

show that CoI classification is not relevant to words presented in documents. Hence, training supervised classifiers from BOW or doc2vec does not give better results compared to using *Stylometric* features only. Adding the *Stylometric* features to BOW or doc2vec offers marginal improvement for SVM or NB, although it has no effect for MaxEnt.

### 5.3.5 One-Class Classification

In Section 5.3.4, we train supervised classifiers from a dataset containing both CoI and non-CoI documents for binary classification. One problem we encountered is that there is **no-degree of assurance** that the items in our non-CoI category are purely non-CoI documents, as they were merely selected randomly from the same Wikipedia categories as CoI articles, with no concrete certainty that they are all non-CoI. Our problem could be potentially solved by one-class classification [146, 260, 115], in which one of the target class is well represented by instances in the training data with little or no other class present. One-class classification has been used and referred to as different concepts depending on application areas such as Outlier Detection, Novelty Detection or Concept Learning [115]. The problem of One-class classification is harder than the problem of conventional classification as a result of the one-sided nature of the dataset. One-class classification makes it difficult to decide which attributes should be used to best separate target and non-target (i.e., CoI and non-CoI in our case).

In [226], adapting SVM to the one-class classification problem has been proposed. Essentially, the input data are first mapped into a high dimensional feature space via a kernel. The origin is considered as the only member of the second class. Then the algorithm iteratively finds the maximal margin hyperplane which best separates the training data from the origin. In our experiments here, we used one-class SVM implementation in the LIBSVM<sup>14</sup> with default parameters.

Table 5.7: CoI detection results using one-class classification.

Feature Sets	Precision	Recall	F-Measure
Stylometric	<b>0.74</b>	<b>0.55</b>	<b>0.63</b>
Presentational	0.69	0.52	0.59
Bias	0.72	0.54	0.62
Emotion	0.73	0.55	0.62
All Features	0.72	0.53	0.61
Best features	0.73	0.54	0.62

Table 5.7 shows the CoI detection results using one-class classification by 10-fold cross-validation

<sup>14</sup><http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

trained on the CoI-related documents only. It can be observed that using *Stylometric* features gives the best results compared to other feature sets although the improvement in F-measure compared to the *Bias* or *Emotion* features is only marginal. We also notice that the precision values, which are in the range of 0.69 to 0.74, are much higher than those achieved based on binary classification where the typical precision values are between 0.58 and 0.64. However, the recall values are lower (0.52~0.55 cf. 0.81~0.94). This shows that if we aim to achieve high recall values for CoI detection, then binary classification should be used. However, if high precision values are more desirable, then one-class classification should be used instead. From our results, when comparing the best F-measure performance between one-class and binary classification, we find that binary classification outperforms one-class classification by 4%, we can say that given that this margin is not very large one-class classification can be considered reasonably effective for CoI detection.

### 5.3.6 Comparison with an Existing Approach to Vandalism Detection

There is no prior approach to content-based CoI detection from Wikipedia. Existing work to bias or vandalism detection often made use of metadata such as anonymity, edit frequency, author reputation, etc., and performed classification at the sentence-level. As we don't have the relevant metadata available and there are no sentence-level annotations in our dataset, directly comparing our approach with existing work is difficult. Nevertheless, we re-implemented an approach proposed in [177] in which their best F-measure and AUC were achieved using LogitBoost and Random Forest, respectively, ranking in the first place of the PAN'10 Wikipedia vandalism detection task [211]. Since we don't have edit histories available, we exclude features relating to edit histories and only extract other stylometric features and features analogous to vulgarism frequency and vulgarism impact and train LogitBoost for 500 iterations. The results in comparison to our best ones are listed in Table 5.8. It can be observed that both our binary and one-class classifiers outperform LogitBoost with the performance gain in F-measure ranging from 6% to 10%.

Table 5.8: Comparison with an existing approach to vandalism detection.

Method	Precision	Recall	F-Measure
LogitBoost [177]	0.56	0.58	0.57
SVM (binary)	<b>0.74</b>	0.55	0.63
SVM (one-class)	0.58	<b>0.81</b>	<b>0.67</b>



### 5.3.7 Discussion

Our finding of the importance of stylometric features confirms our original hypothesis in Section 5.2 that CoI will inherit linguistic and stylometric features from its parent Wikipedia Category. But our hypothesis that presentation relations would affect CoI was not supported by our experimental results. We found that our hypothesis on CoI articles being more subjective holds true based on the experiment results. Also, the hypothesis that CoI articles contain more expressions of implicit and explicit emotions is also supported by our experimental results.

Our feature selection results show that *Blame*, *Praise* and *Polarity\_Score* are discriminative features for the CoI class as they are ranked in the top 3 positions by CFS. However, in binary classification results, using features from the *Emotion* category gives worse results compare to the *Stylometric* category, although it outperforms both *Presentation* and *Bias* categories. The same observation holds for one-class classification. Using *Stylometric* features consistently outperform other feature sets for both binary and one-class classification. Also, it seems that articles with a higher rate of coordinating conjunctions and adjectives per sentence have a higher chance of belonging to the CoI category.

We have also compared our approach with existing work to vandalism detection. The results show that using our defined set of features, we are able to achieve better results. Existing work largely made use of metadata such as anonymity, edit frequency, author reputation, etc. We could consider exploring features extracted from metadata in addition to content-based features for CoI detection in our future work.

In this chapter, we presented our work on blame/praise detection in text based on the “Path Model of Blame”, with the aim of addressing our third research question: *Can the intelligent combination of emotion features help with conflict-of-interest detection in Wikipedia articles?*.

The work presented in this chapter tackles a unique problem for the automatic detection of Conflict of Interest (CoI) articles in Wikipedia entries based on the content of the articles. We have shown that the CoI detection task is a complex problem but with carefully engineered feature sets, it is possible to identify CoI articles with an F-measure of 0.67 using SVM. We have also found that out of four different sets of features, *Stylometric* features help the most with CoI detection. In addition to binary classification, we have experimented with one-class classification and shown that

while binary classification gives higher recall values, one-class classification attains higher precision values.

## Chapter 6

### Conclusion and Future Work

This thesis investigated the problem of emotion detection in text, focusing primarily on implicit emotion detection and application areas for implicit emotion detection. We started by introducing the problem and our motivation for addressing this problem. We explained the role of emotions in everyday life and hence the benefits of our work in various sectors.

Next, we carried out a review of existing work highlighting both the pros and cons of various approaches, explained the methodologies and datasets used in detecting the emotion in text. We proposed a rulebased approach for implicit emotion detection, which can be used for classification in the absence of labelled data. We conducted experiments with three distinct datasets and compared our results with standard baseline supervised classifiers and our approach outperformed the baseline.

We identified a gap in detecting complex emotions such as guilt and remorse. These complex emotions require one to establish blameworthiness or praiseworthiness. We created a dataset for this task, annotating over 7000 comments for blame, praise and fine-grained self-blame, other-blame, self-praise and other-praise. We created a model for blame detection by adapting the psychology Path Model to Blame and conducted experiments with our dataset using our adapted model for blame/praise detection and compared our results with baseline algorithms trained with the same dataset. Our approach performed reasonably well with the fine-grained classification against the baseline.

Since implicit emotion detection might be potentially useful for conflict-of-interest (CoI) detection in Wikipedia articles, we built a CoI corpus by crawling Wikipedia and explored various types of features including linguistic and stylometric features, presentation features, bias features and emotion features for CoI detection. Our results showed that although emotion features are more

important than others when using Naive Bayes, the best performance is obtained by training SVM on linguistic and stylometric features only

## 6.1 Analysis and Evaluation

In chapter 3, we proposed a rule-based approach for implicit emotion detection. We also explore multiple emotion detection from a single sentence and evaluate our work on three different datasets. The work in this chapter addressed our first research sub-question. In chapter 4, we presented our work on detecting expressions of blame in text, presenting a model for blame detection in text. The work in this chapter addresses our second research sub-question. However, we identified various items that may affect the performance of our approach in these two chapters:

- *Error propagated from other NLP tasks*: Our work relies on various natural language processing tasks such as word sense disambiguation, part of speech tagging and many more. These tasks enable our approach to correctly assign values to the various variables used to detect the underlying implicit emotion in text. To this end, if an error exists in any of the NLP tasks, said error will be propagated through the pipeline to the final implicit emotion categorization task. For example, if there is an error in the part of speech tagging task, this error will affect the word sense disambiguation task as well as the tense detection task and eventually the emotion detection task. We try to reduce the overall impact of this error by working on a sentence level (processing whole documents sentence by sentence) and by ensuring that the NLP task approaches used and libraries are up to date and utilising latest research and models.
- *The OCC model emotion detection rules*: Looking at the data in our various datasets, aspects of the OCC model's rules seem to have a reasonable effect on the overall performance of the approach. Such as if a sentence does not contain an *Object* or a *Verb* as with some of the news headlines in the semeval dataset. This inability to detect a *Verb+Object* relation makes detecting an event difficult and in turn the detection of an emotion. The rules can also affect the detection of specific emotions as all the variable parameters for said emotion must be met. For example, an emotion like *Fear* can only be identified if the sentence is identified to be in the PAST or FUTURE tense. The effect of this flaw can however be reduced by identifying additional semantic and syntactic attributes that could be associated with the rules along with a fuzzy matching technique to reduce the rigidness of the rules.

- *Polarity detection dependence*: Although our approach has shown reasonable performance when compared to supervised machine learning approaches, it has a very strong reliance on polarity detection. To mitigate the chances of a polar term not been correctly identified, we decided to use a majority vote approach with 5 lexical resources. However, it remains that failure to correctly detect the polarity of an event, action or sentence can ultimately result in an inability to detect the emotion of the sentence.
- *Morality requirement for detection of blame VS real-world expression of blame*: In chapter 4, we presented our work on blame/praise detection in text based on the “Path Model of Blame”. We identified that for blame to exist that the agent to be blamed must meet a moral requirement, i.e mosquitoes cannot be blamed for malaria or earthquake for the destruction of buildings. However, we find that in our day to day communication people tend to blame such agents for the various outcomes of events. This also means that expressions of blame/praise in the dataset where the agent is not considered to be morally responsible will not be correctly classified as blame/praise. This is primarily because although the model is psychologically correct, real-world expression of blame may not always follow this path.

## 6.2 Contributions

Here we highlight our main contribution as a result of this thesis. In our research, we made use of models from the area of psychology of emotions and combined our understanding of these psychology models with computational linguistics.

Previous work on emotion detection in text relied heavily on labeled data and there was also very little work done on implicit emotion detection. In this thesis, we investigated and created a rule-based approach for implicit emotion detection which does not rely on the availability of labeled data.

Most of the prior work on emotions in text is limited to the available datasets to the point that the limitation in resources has also affected the direction of the research on emotion recognition. In this thesis using the ISEAR dataset, we built our own dataset that is labeled with praise, blame and with blame/praise direction (self or other). We created a model for blame and praise detection in text, an area which to the best of our knowledge is the first of its kind.

We researched the use of emotion features for content-based conflict-of-interest detection in

Wikipedia articles. We found that this area has not been researched extensively and created a dataset which can be used by future researchers in this area. We also found the emotion features influence results in CoI detection but the best features are the linguistic and stylometric features.

### **6.3 Future work**

The current work can be extended in several ways as discussed below:

- Text can trigger emotions of the reader and can also reflect emotions of the writer. Our approaches do not try to distinguish between these two types of emotions. In future, it is important to discriminate explicitly these two emotion types by investigating and utilizing the work of Tang et al. [258] on emotion modeling from writer/reader perspectives and Yang et al. [293] on emotion analysis of social media from writer/reader perspectives. This distinction helps improve the oversimplification of emotion categories problem, as the emotions of the writer may vary from that of the reader. For example, in the health sector, a comment on a health-related forum may arouse emotions of sadness among readers but the writer may be experiencing a deep level of guilt.
- For the Fear emotion, we found that exclamation marks and idioms were most often used as clues for fear. Perhaps looking into the work of Hancock et al [94], they conducted experiments on 40 dyadic interactions and found that users tend to use both verbal strategies such as changes in disagreement, sentence types, the use of affect terms and verbosity, and nonverbal strategies such as the use of punctuation (e.g., exclamation marks) to convey emotion in text.
- We currently do not try to identify ironic and sarcastic sentences. In future it would be appropriate to investigate performance on ironic/sarcastic sentences and help distinguish sentences of such category.
- Our approaches have been largely evaluated on formal text. It would be interesting to see how our approaches perform on informal short text such as tweets and social media posts. We will also improve the identification of emotions involving intensity variables and unexpectedness variables by examining how adverbs and adjectives influence the emotion of sentences. Our approach has not taken into considerations elements of modern informal short text such as

hashtags, emojis and the language specifically in such environments (OMG, LOL and many more).

- Our implicit emotion detection approach performs very poorly with the "Fear" emotion. We believe that by taking a close look at the construction of the fear sentences in the datasets, as well as some of the patterns presented in the work of Fellbaum et al [73] on emotion verb scales, the performance of the algorithm on the "Fear" category can be greatly improved.
- Investigate methods to improve our approach with informal short text such as tweets. We will also improve the identification of emotions involving intensity variables and unexpectedness variables by examining how adverbs and adjectives influence the emotion of sentences (for emotions like "Surprise" and "Shock").
- Investigate possibility to consider a hybrid of the OCC model and other closely related models such as the one proposed in [271] to help improve general performance.
- Explore other types of features extracted from metadata of Wikipedia articles such as editors' information, editing history, article history, associated IP addresses and evaluate their impact on the performance of CoI detection. It is possible that articles in different Wikipedia categories might follow different writing styles (e.g., Wikipedia entries about people and about organisations). One possible direction is to build category-specific classifiers for CoI detection. Finally, to avoid expensive feature engineering, it is possible to learn feature representations and classifiers simultaneously by investigating various deep learning architectures.

## **Appendix A**

### **List of Publications**

Below is the list of publications that contributed to various chapters in this thesis:

1. U. Orizu and Y. He. A Rule-Based Approach to Implicit Emotion Detection in Text, The 20th International Conference on Applications of Natural Language to Information Systems (NLDB), Passau, Germany, June 2015.
2. U. Orizu and Y. He. Detecting Expressions of Blame or Praise in Text, The 10th edition of the Language Resources and Evaluation Conference (LREC), Portoro, Slovenia, May, 2016.
3. U. Orizu and Y. He. Content-Based Conflict of Interest Detection for Articles in Wikipedia, The 11th edition of the Language Resources and Evaluation Conference (LREC), Miyazaki, Japan, May 2018.



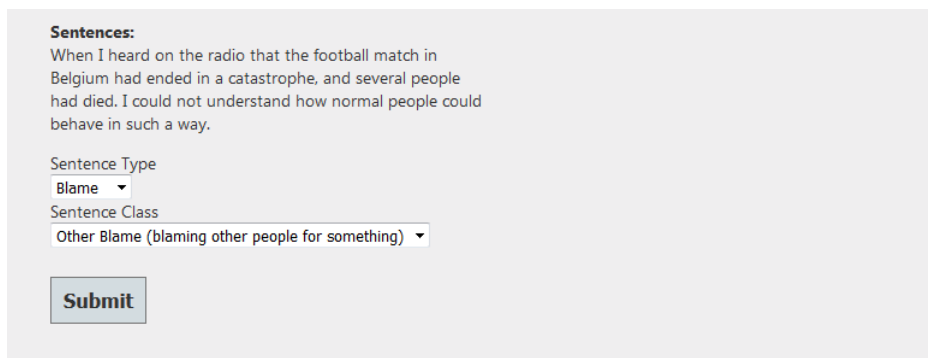
## Appendix B

### Annotation Sample and Guidelines

These are the instructions provide along with sample comments provided to our two annotators for their task.

**Aim:** The following instructions are for the annotation of blame and praise expressed in text, in order to produce a dataset that is to be used in automated evaluation of blame and praise detection algorithms.

Using the web interface provided:



The screenshot shows a web interface for sentence annotation. It features a text area with the following content: "Sentences: When I heard on the radio that the football match in Belgium had ended in a catastrophe, and several people had died. I could not understand how normal people could behave in such a way." Below the text area, there are two dropdown menus. The first is labeled "Sentence Type" and has "Blame" selected. The second is labeled "Sentence Class" and has "Other Blame (blaming other people for something)" selected. A "Submit" button is located at the bottom of the form.

Figure B.1: User Interface sample.

- Read each comment at least twice. The first time, read for overall meaning and impressions. The second time, read more carefully and identify relevant ideas and purpose.
- Begin to annotate.
  - Select from the " Sentence Type " dropdown which category the overall sentence belongs to
  - A second dropdown will appear depending on you previous selections

– Select from the " Sentence Class " dropdown which direction the overall sentence belongs to

- Use the next and previous buttons to move forward and backwards between comments

### **Literary Term Definitions**

**Blame:** feel that (someone or something) is responsible for a fault or wrong. For example: *I feel guilty when when I realize that I consider material things more important than caring for my relatives. I feel very self-centered.*

**Praise:** the expression of approval or admiration for someone or something. For example: *When I was informed that I had been accepted as a student of Psychology.*

**Other:** when a sentence does not fall into praise or blame category. For example: *When I left New York, and all my family and my friends behind me.*

**Self Blame:** blaming one's self for something

**Other Blame:** blaming other people for something

**Self Praise:** praising one's self for something

**Other Praise:** praising other people for something

## Appendix C

### List of Intensifiers

#### **word, category, weight**

*absolutely,booster,0.5*

*altogether,booster,0.5*

*blind,booster,0.5*

*clean,booster,0.5*

*completely,booster,0.5*

*dead,booster,0.5*

*downright,booster,0.5*

*entirely,booster,0.5*

*fast,booster,0.5*

*full,booster,0.5*

*fully,booster,0.5*

*outright,booster,0.5*

*perfectly,booster,0.5*

*plain,booster,0.5*

*quite,booster,0.5*

*stark,booster,0.5*

*thoroughly,booster,0.5*

*totally,booster,0.5*

*utterly,booster,0.5*

*wholly,booster,0.5*

*wide,booster,0.5*  
*awfully,booster,0.5*  
*damn,booster,0.5*  
*damned,booster,0.5*  
*deeply,booster,0.5*  
*truly,booster,0.5*  
*genuinely,booster,0.5*  
*really,booster,0.5*  
*dreadfully,booster,0.5*  
*enormously,booster,0.5*  
*exceedingly,booster,0.5*  
*extremely,booster,0.5*  
*frightfully,booster,0.5*  
*greatly,booster,0.5*  
*heavily,booster,0.5*  
*highly,booster,0.5*  
*horribly,booster,0.5*  
*immensely,booster,0.5*  
*incredibly,booster,0.5*  
*infinitely,booster,0.5*  
*jolly,booster,0.5*  
*remarkably,booster,0.5*  
*so,booster,0.5*  
*strongly,booster,0.5*  
*terribly,booster,0.5*  
*tremendously,booster,0.5*  
*very,booster,0.5*  
*well,booster,0.5*  
*all-fired,booster,0.5*  
*bloody,booster,0.5*

*damn,booster;0.5*  
*mighty,booster;0.5*  
*powerful,booster;0.5*  
*mightily,booster;0.5*  
*right,booster;0.5*  
*in truth,booster;0.5*  
*deucedly,booster;0.5*  
*insanely,booster;0.5*  
*deadly,booster;0.5*  
*madly,booster;0.5*  
*devilishly,booster;0.5*  
*literally,booster;0.5*  
*candidly,booster;0.5*  
*candid,booster;0.5*  
*scoldingly,booster;0.5*  
*frankly,booster;0.5*  
*always,booster;0.5*  
*honestly,booster;0.5*  
*generally,booster;0.5*  
*marvellously,booster;0.5*  
*marvelously,booster;0.5*  
*superbly,booster;0.5*  
*terrifically,booster;0.5*  
*toppingly,booster;0.5*  
*wonderfully,booster;0.5*  
*wondrous,booster;0.5*  
*wondrously,booster;0.5*  
*goddam,booster;0.5*  
*goddamn,booster;0.5*  
*goddamned,booster;0.5*

*hellishly,booster;0.5*  
*infernally,booster;0.5*  
*about,downtoner;-0.5*  
*almost,downtoner;-0.5*  
*most,downtoner;-0.5*  
*nearly,downtoner;-0.5*  
*near,downtoner;-0.5*  
*nigh,downtoner;-0.5*  
*virtually,downtoner;-0.5*  
*well-nigh,downtoner;-0.5*  
*approximately,downtoner;-0.5*  
*about, downtoner;-0.5*  
*close to, downtoner;-0.5*  
*just about, downtoner;-0.5*  
*some, downtoner;-0.5*  
*roughly, downtoner;-0.5*  
*more or less, downtoner;-0.5*  
*around,downtoner;-0.5*  
*or so ,downtoner;-0.5*  
*slightly,downtoner;-0.5*  
*somewhat,downtoner;-0.5*  
*partially, downtoner;-0.5*  
*partly, downtoner;-0.5*  
*part, downtoner;-0.5*  
*in part,downtoner;-0.5*  
*barely, downtoner;-0.5*  
*hardly, downtoner;-0.5*  
*just, downtoner;-0.5*  
*scarcely, downtoner;-0.5*  
*scarce downtoner;-0.5*



## References

- [1] B Thomas Adler and Luca De Alfaro. A content-driven reputation system for the wikipedia. In *Proceedings of the 16th international conference on World Wide Web*, pages 261–270. ACM, 2007.
- [2] B Thomas Adler, Luca De Alfaro, Santiago M Mola-Velasco, Paolo Rosso, and Andrew G West. Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 277–288. Springer, 2011.
- [3] Rodrigo Agerri and Ana Garca-Serrano. Q-wordnet: Extracting polarity from wordnet senses. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
- [4] Charu C Aggarwal and ChengXiang Zhai. *Mining text data*. Springer Science & Business Media, 2012.
- [5] Ameeta Agrawal and Aijun An. Unsupervised emotion detection from text using semantic and syntactic relations. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*, volume 1, pages 346–353. IEEE, 2012.
- [6] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 579–586. Association for Computational Linguistics, 2005.



- [7] Omar Alonso and Kyle Shiells. Timelines as summaries of popular scheduled events. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 1037–1044. International World Wide Web Conferences Steering Committee, 2013.
- [8] Bengt Altenberg. Amplifier collocations in spoken english. *English Computer Corpora: Selected Papers and Research Guide/Ed. S. Johansson and A.-B. Stenstrom. Berlin: Mouton*, pages 127–147, 1991.
- [9] Saima Aman. *Recognizing emotions in text*. PhD thesis, University of Ottawa, 2007.
- [10] Saima Aman and Stan Szpakowicz. Identifying expressions of emotion in text. In *Text, Speech and Dialogue*, pages 196–205. Springer, 2007.
- [11] Scott Anderson. Coercion. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, spring 2014 edition, 2014.
- [12] Shlomo Argamon, Moshe Koppel, James W Pennebaker, and Jonathan Schler. Automatically profiling the author of an anonymous text. *Communications of the ACM*, 52(2): pages119–123, 2009.
- [13] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4): pages 555–596, 2008.
- [14] M Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, and Fazal Masud Kundi. A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, 4(3): pages181–186, 2014.
- [15] Pablo Azar. Sentiment analysis in financial news. *Harvard University*, 2009.
- [16] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.
- [17] Collin F Baker, Charles J Fillmore, and John B Lowe. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*

- and *17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.
- [18] Iris Bakker, Theo van der Voordt, Peter Vink, and Jan de Boon. Pleasure, arousal, dominance: Mehrabian and russell revisited. *Current Psychology*, 33(3): pages 405–421, 2014.
- [19] RC Balabantaray, Mudasir Mohammad, and Nibha Sharma. Multi-class twitter emotion classification: A new approach. *International Journal of Applied Information Systems*, 4(1): pages 48–53, 2012.
- [20] Alexandra Balahur and Jesus M Hermida. Extending the emotinet knowledge base to improve the automatic detection of implicitly expressed emotions from text.
- [21] Alexandra Balahur, Jesús M Hermida, and Andrés Montoyo. Detecting implicit expressions of sentiment in text based on commonsense knowledge. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 53–60. Association for Computational Linguistics, 2011.
- [22] Alexandra Balahur, Jesus M. Hermida, and Andres Montoyo. Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53(4): pages 742 –753, 2012. 1) Computational Approaches to Subjectivity and Sentiment Analysis 2) Service Science in Information Systems Research : Special Issue on {PACIS} 2010.
- [23] Alexandra Balahur, Jesús M Hermida, and Andrés Montoyo. Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53(4): pages 742–753, 2012.
- [24] Alexandra Balahur, Jesus M Hermida, Andres Montoyo, and Rafael Munoz. Emotinet: A knowledge base for emotion detection in text built on the appraisal theories. In *Natural Language Processing and Information Systems*, pages 27–39. Springer, 2011.
- [25] Alexandra Balahur, Jesus M. Hermida, Andres Montoyo, and Rafael Munoz. Detecting implicit expressions of affect in text using emotinet and its extensions. *Data and Knowledge Engineering*, 88(0): pages 113 – 125, 2013.

- [26] Alexandra Balahur, Jesús M Hermida, and Hristo Tanev. Detecting implicit emotion expressions from text using ontological resources and lexical learning. In *New Trends of Research in Ontologies and Lexical Resources*, pages 235–255. Springer, 2013.
- [27] Alexandra Balahur, JesusM. Hermida, and Hristo Tanev. Detecting implicit emotion expressions from text using ontological resources and lexical learning. In Alessandro Oltramari, Piek Vossen, Lu Qin, and Eduard Hovy, editors, *New Trends of Research in Ontologies and Lexical Resources*, Theory and Applications of Natural Language Processing, pages 235–255. Springer Berlin Heidelberg, 2013.
- [28] Alexandra Balahur, JesusM. Hermida, and Hristo Tanev. Detecting implicit emotion expressions from text using ontological resources and lexical learning. In Alessandro Oltramari, Piek Vossen, Lu Qin, and Eduard Hovy, editors, *New Trends of Research in Ontologies and Lexical Resources*, Theory and Applications of Natural Language Processing, pages 235–255. Springer Berlin Heidelberg, 2013.
- [29] Alexandra Balahur, Ralf Steinberger, Mijail Kadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*, 2013.
- [30] Eugene Yuta Bann. Discovering basic emotion sets via semantic clustering on a twitter corpus. *arXiv preprint arXiv:1212.6527*, 2012.
- [31] Charles Bazerman. *The informed writer*. Houghton Mifflin Company, 1981.
- [32] Núria Bel, Sheila Queralt, Maria Spassova, and M Teresa Turell. The use of sequences of linguistic categories in forensic written text comparison revisited. In *Proceedings of The International Association of Forensic Linguists Tenth Biennial Conference*, pages 192–209. Centre for Forensic Linguistics, Aston University, 2012.
- [33] Jerome R Bellegarda. Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 1–9. Association for Computational Linguistics, 2010.

- [34] Shruti Bhosale, Heath Vinicombe, and Raymond J Mooney. Detecting promotional content in wikipedia. In *EMNLP*, pages 1851–1857, 2013.
- [35] Serajul I Bhuiyan. Social media and its effectiveness in the political reform movement in egypt. *Middle East Media Educator*, 1(1): pages 14–20, 2011.
- [36] Robert I Binnick. *Time and the verb: A guide to tense and aspect*. Oxford University Press, 1991.
- [37] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3: pages 993–1022, 2003.
- [38] Alice Brand. Social cognition, emotions, and the psychology of writing. *Journal of Advanced Composition*, pages 395–407, 1991.
- [39] Alice G Brand. Hot cognition: Emotions and writing behavior. *Journal of Advanced Composition*, pages 5–15, 1985.
- [40] Alice G. Brand. The why of cognition: Emotion and the writing process. *College Composition and Communication*, 38(4): pages 436–443, 1987.
- [41] Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. Combining strengths, emotions and polarities for boosting twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 2. ACM, 2013.
- [42] Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 804–812. Association for Computational Linguistics, 2010.
- [43] John A Bullinaria and Joseph P Levy. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods*, 39(3): pages 510–526, 2007.

- [44] Joan L Bybee and Östen Dahl. *The creation of tense and aspect systems in the languages of the world*. John Benjamins Amsterdam, 1989.
- [45] Ewa S Callahan and Susan C Herring. Cultural bias in wikipedia content on famous persons. *Journal of the Association for Information Science and Technology*, 62(10): pages 1899–1915, 2011.
- [46] Rafael A Calvo and Sunghwan Mac Kim. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3): pages 527–543, 2013.
- [47] Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2): pages 15–21, 2013.
- [48] Daniel Cer, Marie-Catherine de Marneffe, Daniel Jurafsky, and Christopher D. Manning. Parsing to stanford dependencies: Trade-offs between speed and accuracy. In *7th International Conference on Language Resources and Evaluation (LREC 2010)*, 2010.
- [49] Deepayan Chakrabarti and Kunal Punera. Event summarization using tweets. In *The International Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [50] Rishi Chandy. Wikiganda: Identifying propaganda through text analysis. *Caltech Undergraduate Research Journal*. Winter, 2009: pages 6–11, 2008.
- [51] Rishi Chandy. Wikiganda: Identifying propaganda through text analysis. *Caltech Undergraduate Research Journal*. Winter, 2009, 2008.
- [52] Yi Chang, Xuanhui Wang, Qiaozhu Mei, and Yan Liu. Towards twitter context summarization with user influence models. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 527–536. ACM, 2013.
- [53] Si-Chi Chin, W Nick Street, Padmini Srinivasan, and David Eichmann. Detecting wikipedia vandalism with active learning and statistical language models. In *Proceedings of the 4th workshop on Information credibility*, pages 3–10. ACM, 2010.
- [54] Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In *EMNLP*, volume 2004, pages 33–40, 2004.

- [55] Freddy Chong Tat Chua and Sitaram Asur. Automatic summarization of events from social media. *The 7th International Conference on Weblogs and Social Media*, pages 81–90, 2013.
- [56] Gerald L Clore and Andrew Ortony. Psychological construction in the occ model of emotion. *Emotion Review*, 5(4): pages 335–343, 2013.
- [57] Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.
- [58] Sean Coughlan. Spelling mistakes 'cost millions' in lost online sales, 2011 (accessed February 21, 2014). Available at <http://www.bbc.co.uk/news/education-14130854>.
- [59] Maral Dadvar, Claudia Hauff, and FMG De Jong. Scope of negation detection in sentiment analysis. pages 16–20, 2011.
- [60] Om P Damani. Improving pointwise mutual information (pmi) by incorporating significant co-occurrence. *arXiv preprint arXiv:1307.0596*, 2013.
- [61] Antonio Damasio. *Descartes error: Emotion, reason and the human brain*. Penguin Group, 1994.
- [62] Antonio Damasio. *Descartes' error: Emotion, reason and the human brain*. Random House, 2008.
- [63] Taner Danisman and Adil Alpkocak. Feeler: Emotion classification of text using vector space model. In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, volume 1, page 53, 2008.
- [64] Dipankar Das and Sivaji Bandyopadhyay. Sentence level emotion tagging. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.
- [65] Bart Desmet and Véronique Hoste. Emotion detection in suicide notes. *Expert Systems with Applications*, 40(16): pages 6351–6358, 2013.
- [66] Dictionary. Cambridge dictionary, 2010.

- [67] Harris Drucker, Chris JC Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9: pages 155–161, 1997.
- [68] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4): pages 169–200, 1992.
- [69] Andrew Eshleman. Moral responsibility. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2014 edition, 2014.
- [70] Andrea Esuli and Fabrizio Sebastiani. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 617–624. ACM, 2005.
- [71] Yehuda N. Falk. The english auxiliary system: A lexical-functional analysis. *Language*, 60(3):483–509, 1984.
- [72] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014.
- [73] Christiane Fellbaum and Yvette Yannick Mathieu. A corpus-based construction of emotion verb scales. *Linguistic Approaches to Emotions in Context*, 241:99, 2014.
- [74] Joseph L Fleiss, Jacob Cohen, and BS Everitt. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5):323, 1969.
- [75] Jay Friedenberg and Gordon Silverman. *Cognitive science: an introduction to the study of mind*. Sage, 2011.
- [76] Susan R Fussell. *The verbal communication of emotions: interdisciplinary perspectives*. Psychology Press, 2002.
- [77] Yannick Gaspar and Luis Macedo. Towards building user affect models from tweets: a comparative study with various datasets, features and algorithms. 2012.
- [78] Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. Hierarchical approach to emotion recognition and classification in texts. In *Advances in Artificial Intelligence*, pages 40–50. Springer, 2010.

- [79] A Gill, R French, Darren Gergle, and Jon Oberlander. Identifying emotional characteristics from short blog texts. In *Proc. 30th ANN. Conf. Cognitive Science Soc., BC Love, K. McRae, and VM Sloutsky, eds*, pages 2237–2242, 2008.
- [80] Alastair J Gill, Robert M French, Darren Gergle, and Jon Oberlander. The language of emotion in short blog texts. In *CSCW*, volume 8, pages 299–302, 2008.
- [81] Alastair J Gill, Darren Gergle, Robert M French, and Jon Oberlander. Emotion rating from short blog texts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1121–1124. ACM, 2008.
- [82] Alastair J Gill, Jon Oberlander, and Jonathan I Ellis. Texttone: Emotion detection in short texts. Pilot project report, School of Informatics, University of Edinburgh 2005.
- [83] Roxana Girju. Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*, pages 76–83. Association for Computational Linguistics, 2003.
- [84] Alfio Gliozzo and Carlo Strapparava. Domain kernels for text categorization. In *Proceedings of the Ninth Conference on Computational Natural Language Learning, CONLL '05*, pages 56–63, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
- [85] Annette M Green. Kappa statistics for multiple raters using categorical classifications. In *Proceedings of the 22nd annual SAS User Group International conference*, volume 2, page 4. San Diego, 1997.
- [86] Yi Guo, Nan Hua, and Zhiqing Shao. An algorithm for detection of cognitive intentionality in text analysis. In *Information Technology and Computer Science, 2009. ITCS 2009. International Conference on*, volume 2, pages 176–179. IEEE, 2009.
- [87] Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. Emotion detection in email customer care. *Computational Intelligence*, 29(3): pages 489–505, 2013.
- [88] Diana Hacker. *The Bedford handbook for writers*. Bedford Books of St. Martin’s Press, 1991.



- [89] Mohamed H Haggag. Frame semantics evolutionary model for emotion detection. *Computer and Information Science*, 7(1): page136, 2014.
- [90] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1): pages10–18, 2009.
- [91] Mentor Hamiti, Arsim Susuri, and Agni Dika. Machine learning and the detection of anomalies in wikipedia. In *PDF). Proceedings of the 19th International Conference on Circuits, Systems, Communications and Computers*, 2012.
- [92] Mentor Hamiti, Arsim Susuri, and Agni Dika. Machine learning and the detection of anomalies in wikipedia. In *Recent Advances in Communications, Proceedings of the 19th International Conference on Communications, Zakynthos Island, Greece*, 2015.
- [93] Jeffrey T Hancock, Kailyn Gee, Kevin Ciaccio, and Jennifer Mae-Hwah Lin. I’m sad you’re sad: emotional contagion in cmc. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 295–298. ACM, 2008.
- [94] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 929–932. ACM, 2007.
- [95] Lars Kai Hansen, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni, and Michael Etter. Good friends, bad news-affect and virality in twitter. *Future information technology*, pages 34–43, 2011.
- [96] Manoj Harpalani, Michael Hart, Sandesh Singh, Rob Johnson, and Yejin Choi. Language of vandalism: Improving wikipedia vandalism detection via stylometric analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 83–88. Association for Computational Linguistics, 2011.
- [97] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.

- [98] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [99] Pamela Hieronymi. The force and fairness of blame. *Philosophical Perspectives*, 18(1): pages115–148, 2004.
- [100] Pamela Hieronymi. Shers defense of blame. *Philosophical Studies*, 137(1): pages19–30, 2008.
- [101] Dan Hill. *Emotionomics: Winning hearts and minds*. Adams Business & Professional, 2007.
- [102] Dion Hoe Lian Goh and Chei Sian Lee. An analysis of tweets in response to the death of michael jackson. *Aslib Proceedings*, 63(5): pages 432–444, 2011.
- [103] Alexander Hogenboom, Paul Van Iterson, Bas Heerschop, Flavius Frasinca, and Uzay Kaymak. Determining negation scope and strength in sentiment analysis. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 2589–2594. IEEE, 2011.
- [104] Philip N Howard and Muzammil M Hussain. The role of digital media. *Journal of Democracy*, 22(3):35–48, 2011.
- [105] Luwen Huangfu, Wenji Mao, Daniel Zeng, and Lei Wang. Occ model-based emotion extraction from online reviews. In *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on*, pages 116–121. IEEE, 2013.
- [106] Nitin Indurkha and Fred J Damerau. *Handbook of natural language processing*, volume 2. CRC Press, 2010.
- [107] Diana Inkpen, Fazel Keshtkar, and Diman Ghazi. Analysis and generation of emotion in texts. *KEPT*, pages 3–13, 2009.
- [108] Kelly Y Itakura and Charles LA Clarke. Using dynamic markov compression to detect vandalism in the wikipedia. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 822–823. ACM, 2009.
- [109] Sara Javanmardi, David W McDonald, and Cristina V Lopes. Vandalism detection in wikipedia: a high-performing, feature-rich model and its reduction through lasso. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 82–90. ACM, 2011.

- [110] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.
- [111] Dan Jurafsky and James H Martin. *Speech and language processing*, volume 3. Pearson London, 2014.
- [112] EC-C Kao, Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, and Von-Wun Soo. Towards text-based emotion detection a survey and possible improvements. In *Information Management and Engineering, 2009. ICIME 09. International Conference on*, pages 70–74. IEEE, 2009.
- [113] Aamera ZH Khan, Mohammad Atique, and VM Thakare. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE)*, page 89, 2015.
- [114] Muhammad Asif Hossain Khan, Danushka Bollegala, Guangwen Liu, and Kaoru Sezaki. Delineating real-time events by identifying relevant tweets with popular discussion points. *HUMAN*, 2(3): page136, 2013.
- [115] Shehroz S Khan and Michael G Madden. A survey of recent trends in one class classification. In *Irish Conference on Artificial Intelligence and Cognitive Science*, pages 188–197. Springer, 2009.
- [116] Seungyeon Kim, Fuxin Li, Guy Lebanon, and Irfan Essa. Beyond sentiment: The manifold of human emotions. *arXiv preprint arXiv:1202.1568*, 2012.
- [117] Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A Calvo. Evaluation of unsupervised emotion models to textual affect recognition. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 62–70. Association for Computational Linguistics, 2010.
- [118] Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. Extending verbnet with novel verb classes. In *Proceedings of LREC*, volume 2006, page 1. Citeseer, 2006.

- [119] Serkan Kiranyaz, Turker Ince, Alper Yildirim, and Moncef Gabbouj. Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. *Neural Networks*, 22(10): pages 1448 – 1462, 2009.
- [120] Judith Klavans and Min-Yen Kan. Role of verbs in document analysis. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 680–686. Association for Computational Linguistics, 1998.
- [121] Jacob Kohen. A coefficient of agreement for nominal scale. *Educ Psychol Meas*, 20: pages 37–46, 1960.
- [122] Agata Kolakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michal R Wrobel. Emotion recognition and its application in software engineering. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 532–539. IEEE, 2013.
- [123] Sumedha Kshirsagar. A multilayer personality model. In *Proceedings of the 2Nd International Symposium on Smart Graphics, SMARTGRAPH '02*, pages 107–115, New York, NY, USA, 2002. ACM.
- [124] Florian Kunneman, Christine Liebrecht, and Antal van den Bosch. The (un) predictability of emotional hashtags in twitter. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL*, pages 26–34, 2014.
- [125] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
- [126] David A Lagnado and Shelley Channon. Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108(3): pages 754–770, 2008.
- [127] Brent Lance and Stacy Marsella. Glances, glares, and glowering: How should a virtual human express emotion through gaze? 20: pages 50–69, May 2010.
- [128] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.
- [129] Richard S Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

- [130] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196, 2014.
- [131] Chul Min Lee, Shrikanth S Narayanan, and Roberto Pieraccini. Combining acoustic and language information for emotion recognition. In *Seventh International Conference on Spoken Language Processing, INTERSPEECH, 2002*.
- [132] Jingsheng Lei, Yanghui Rao, Qing Li, Xiaojun Quan, and Liu Wenying. Towards building a social emotion detection system for online news. *Future Generation Computer Systems*, 37: pages 438–448, 2014.
- [133] K Ming Leung. Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, 2007.
- [134] Omer Levy, Yoav Goldberg, and Ido Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3: pages 211–225, 2015.
- [135] Michael Lewis, Jeannette M Haviland-Jones, and Lisa Feldman Barrett. *Handbook of emotions*. Guilford Press, 2010.
- [136] Junhui Li, Guodong Zhou, Hongling Wang, and Qiaoming Zhu. Learning the scope of negation via shallow semantic parsing. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 671–679. Association for Computational Linguistics, 2010.
- [137] Weiyuan Li and Hua Xu. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41(4):1742–1749, 2014.
- [138] Chen Lin, Chun Lin, Jingxuan Li, Dingding Wang, Yang Chen, and Tao Li. Generating event storylines from microblogs. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 175–184. ACM, 2012.
- [139] Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2(4): pages 285–318, 1988.

- [140] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.
- [141] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. Springer, 2012.
- [142] Hugo Liu, Henry Lieberman, and Ted Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132. ACM, 2003.
- [143] Zitao Liu, Wenchao Yu, Wei Chen, Shuran Wang, and Fengyi Wu. Short text feature selection for micro-blog mining. In *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, pages 1–4. IEEE, 2010.
- [144] Steffen Lohmann, Michael Burch, Hansjörg Schmauder, and Daniel Weiskopf. Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 753–756. ACM, 2012.
- [145] Bertram F Malle, Steve Guglielmo, and Andrew E Monroe. A theory of blame. *Psychological Inquiry*, 25(2): pages 147–186, 2014.
- [146] Larry M Manevitz and Malik Yousef. One-class svms for document classification. *Journal of Machine Learning Research*, 2(Dec): pages 139–154, 2001.
- [147] Inderjeet Mani and George Wilson. Robust temporal processing of news. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 69–76. Association for Computational Linguistics, 2000.
- [148] William C Mann and Sandra A Thompson. *Rhetorical structure theory: A theory of text organization*. University of Southern California, Information Sciences Institute, 1987.
- [149] William C Mann and Sandra A Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3): pages 243–281, 1988.

- [150] W. Mao, X. Li, and A. Ge. From causal scenarios to social causality: An attributional approach. *IEEE Intelligent Systems*, 26: pages 48–57, 09 2011.
- [151] Wenji Mao, Ansheng Ge, and Xiaochen Li. From causal scenarios to social causality: An attributional approach. *IEEE Intelligent Systems*, 26(6): pages 48–57, 2011.
- [152] Wenji Mao and Jonathan Gratch. Social causality and responsibility: Modeling and evaluation. In *Intelligent Virtual Agents*, pages 191–204. Springer, 2005.
- [153] Rouse Margaret. Microblogging, January 2009 (accessed February 10, 2014). Available at <http://searchmobilecomputing.techtarget.com/definition/microblogging>.
- [154] Stacy Marsella, Jonathan Gratch, and Paolo Petta. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, pages 21–46, 2010.
- [155] FJAP Mattosinho. PhD thesis: Mining product opinions and reviews on the web. *Technische Universitat Dresden*, 2010. Available at <https://bit.ly/2MEoW1h>
- [156] John D Mayer, Peter Salovey, and David R Caruso. Emotional intelligence: Theory, findings, and implications. *Psychological inquiry*, 15(3): pages197–215, 2004.
- [157] Diana Maynard, Kalina Bontcheva, and Dominic Rout. Challenges in developing opinion mining tools for social media. *Proceedings of@ NLP can u tag# usergeneratedcontent*, 2012, Available at <https://bit.ly/2woZ0wz>.
- [158] Michael McKenna. *Conversation & Responsibility*. OUP USA, 2012.
- [159] Michael McKenna. Directed blame and conversation. *Blame: Its nature and norms*, Oxford University Press pages 119–140, 2012.
- [160] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [161] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

- [162] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11): pages 39–41, 1995.
- [163] Gilad Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, volume 19, 2005.
- [164] Tom M Mitchell. Logistic regression. *Machine learning*, 10:701, 2005.
- [165] Saif Mohammad. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114. Association for Computational Linguistics, 2011.
- [166] Saif Mohammad. Portable features for classifying emotional text. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 587–591. Association for Computational Linguistics, 2012.
- [167] Saif M Mohammad. # emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 246–255. Association for Computational Linguistics, 2012.
- [168] Saif M Mohammad and Svetlana Kiritchenko. Using nuances of emotion to identify personality. *arXiv preprint arXiv:1309.6352*, 2013.
- [169] Saif M Mohammad and Svetlana Kiritchenko. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2): pages 301–326, 2014.
- [170] Saif M Mohammad and Peter D Turney. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34. Association for Computational Linguistics, 2010.
- [171] Saif M Mohammad and Peter D Turney. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3): pages 436–465, 2013.



- [172] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29(3): pages 436–465, Wiley Online Library, 2013.
- [173] Saif M. Mohammad and Peter D. Turney. NRC emotion Lexicon: Crowdsourcing a word-emotion association lexicon. 29(3): pages 416–423, National Research Council, Ca. 2014.
- [174] Saif M Mohammad and Tony Wenda Yang. Tracking sentiment in mail: how genders differ on emotional axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011)*, pages 70–79, 2011.
- [175] Mitra Mohtarami, Man Lan, and Chew Lim Tan. Probabilistic sense sentiment similarity through hidden emotions. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics 1*: pages 983–992, 2013.
- [176] Santiago M Mola-Velasco. Wikipedia vandalism detection. In *Proceedings of the 20th international conference companion on World wide web*, pages 391–396. ACM, 2011.
- [177] Santiago M Mola-Velasco. Wikipedia vandalism detection through machine learning: Feature review and new proposals: Lab report for pan at clef 2010. *arXiv preprint arXiv:1210.5560*, 2012.
- [178] Santiago Moisés Mola-Velasco. Wikipedia vandalism detection through machine learning: Feature review and new proposals: Lab report for PAN at CLEF 2010. *CoRR*, abs/1210.5560, 2012.
- [179] Sara Ahmed Morsy. Recognizing contextual valence shifters in document-level sentiment classification. School Of Sciences & Engineering, American University, Cairo 2011.
- [180] Kaitlyn Mulcrone. Detecting emotion in text. *University of Minnesota–Morris CS Senior Semminar Paper*, 2012.
- [181] Ronald Alan Nash. Cognitive theories of emotion. *Nous*, 23(4): pages 481–504, 1989.
- [182] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. Compositionality principle in recognition of fine-grained emotions from text. In *The International Conference on Weblogs and Social Media (ICWSM)*, 2009.

- [183] ALENA NEVIAROUSKAYA, HELMUT PRENDINGER, and MITSURU ISHIZUKA. Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17: pages 95–135, 1 2011.
- [184] F. Å. Nielsen, M. Etter, and L. K. Hansen. Real-time monitoring of sentiment in business related wikipedia articles, Technical University of Denmark 2013. <https://bit.ly/2wwH421>
- [185] Finn Årup Nielsen. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. *CoRR*, abs/1103.2903, 2011.
- [186] Brendan OConnor, Michel Krieger, and David Ahn. Tweetmotif: Exploratory search and topic summarization for twitter. In *The International Conference on Weblogs and Social Media (ICWSM)*, 2010.
- [187] Aileen Oeberst, Iassen Halatchliyski, Joachim Kimmerle, and Ulrike Cress. Knowledge construction in wikipedia: A systemic-constructivist analysis. *Journal of the Learning Sciences*, 23(2): pages149–176, 2014.
- [188] John Olsson and June Luchjenbroers. *Forensic linguistics*. A&C Black, 2013.
- [189] Udochukwu Orizu and Yulan He. Detecting expressions of blame or praise in text. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may 2016. European Language Resources Association (ELRA).
- [190] Andrew Ortony. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [192] Andrew Ortony. On making believable emotional agents believable. *Trappl et al.(Eds.)(2002)*, pages 189–211, 2002.

- [193] Woodley Packard, Emily M Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan. Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *ACL (1)*, pages 69–78, 2014.
- [194] Martha Palmer. Consistent criteria for sense distinctions. *Computers and the Humanities*, 34(1-2): pages 217–222, 2000.
- [195] Polina Panicheva, John Cardiff, and Paolo Rosso. Personal sense and idiolect: combining authorship attribution and opinion analysis. Dublin Institute of Technology, LREC 2010.
- [196] Manoj Parameswaran and Andrew B Whinston. Social computing: An overview. *Communications of the Association for Information Systems*, 19(1):37, 2007.
- [197] Deepa Paranjpe. Learning document aboutness from implicit user feedback and document structure. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 365–374. ACM, 2009.
- [198] W Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001.
- [199] SS Patil and CG Patil. Study and review of various emotion classifications from text techniques. *International Magazine on Advances in Computer Science and Telecommunications*, 4(1): pages 5–8, 2013.
- [200] Ellie Pavlick, Pushpendre Rastogi, Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. Ppdb 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 425–430, Beijing, China, July 2015. Association for Computational Linguistics.
- [201] James W Pennebaker and Laura A King. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296, 1999.
- [202] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

- [203] Jacob Perkins. *Python 3 Text Processing with NLTK 3 Cookbook*. Packt Publishing Ltd, 2014.
- [204] Rosalind W Picard. *Affective computing*, volume 59. Elsevier, 2000.
- [205] Robert Plutchik. The nature of emotions. *American Scientist*, 89(4): pages 344–350, 2001.
- [206] Livia Polanyi and Annie Zaenen. Contextual valence shifters. *Computing Attitude and Affect in Text*, 20: pages 1–10, 2006.
- [207] Ana-Maria Popescu and Marco Pennacchiotti. Detecting controversial events from twitter. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1873–1876. ACM, 2010.
- [208] Ana-Maria Popescu, Marco Pennacchiotti, and Deepa Paranjpe. Extracting events and event descriptions from twitter. In *Proceedings of the 20th international conference companion on World wide web*, pages 105–106. ACM, 2011.
- [209] Soujanya Poria, Alexander Gelbukh, Dipankar Das, and Sivaji Bandyopadhyay. Fuzzy clustering for semi-supervised learning—case study: Construction of an emotion lexicon. In *Advances in Artificial Intelligence*, pages 73–86. Springer, 2013.
- [210] Martin Potthast, Benno Stein, and Robert Gerling. Automatic vandalism detection in wikipedia. In *European Conference on Information Retrieval*, pages 663–668. Springer, 2008.
- [211] Martin Potthast, Benno Stein, and Teresa Holfeld. Overview of the 1st international competition on wikipedia vandalism detection. In *CLEF (Notebook Papers/LABs/Workshops)*, 2010.
- [212] Changqin Quan and Fuji Ren. Sentence emotion analysis and recognition based on emotion words using ren-cecps. *International Journal of Advanced Intelligence*, 2(1):105–117, 2010.
- [213] Rationality. *The Cambridge Dictionary*. Cambridge University Press, 1999.
- [214] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *ACL (1)*, pages 1650–1659, 2013.

- [215] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proceedings of ACL*, 2013.
- [216] T Raghunadha Reddy, B Vishnu Vardhan, and P Vijayapal Reddy. A survey on authorship profiling techniques. *International Journal of Applied Engineering Research*, 11(5): pages 3092–3102, 2016.
- [217] Andi Rexha, Stefan Klampfl, Mark Kröll, and Roman Kern. Towards authorship attribution for bibliometrics using stylometric features. In *Proc. of the Workshop Mining Scientific Papers: Computational Linguistics and Bibliometrics, 15th International Society of Scientometrics and Informetrics Conference (ISSI), Istanbul, Turkey: <http://ceur-ws.org>*, 2015.
- [218] Laurel Richardson. *Writing strategies: Reaching diverse audiences*, volume 21. Sage, 1990.
- [219] Martin Ringsquandl and Dusan Petkovic. Analyzing political sentiment on twitter. In *Proceedings of AAAI conference*, 2013.
- [220] Kirk Roberts and Sanda M Harabagiu. Statistical and similarity methods for classifying emotion in suicide notes. *Biomedical informatics* pages 195–204, SAGE Publications 2012.
- [221] Kirk Roberts, Michael A Roach, Joseph Johnson, Josh Guthrie, and Sanda M Harabagiu. Empatweet: Annotating and detecting emotions on twitter. In *LREC*, pages 3806–3813, 2012.
- [222] Saulo A. Santos, Aêuser J. S. Valença, and Aarmelo J. A. Bastos-Filho. *Antelligent Aata Engineering and Automated Aearning - ADEAL 2012: 13th Anternational Aonference, Natal, Brazil, August 29-31, 2012. Proceedings*, Achapter Aomparing Particle Swarm Aptimization Approaches Aor Training Aulti-Layer Perceptron Neural Networks Aor Aorecasting, pages 344–351. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [223] J Savoy. Ir multilingual resources at unine, 2011 (accessed JUNE 21, 2016). <http://members.unine.ch/jacques.savoy/clef/index.html>.
- [224] Robert E Schapire. A brief introduction to boosting. In *International Joint Conference on Artificial Intelligence*, volume 99, pages 1401–1406, 1999.

- [225] Robert E Schapire and Yoram Singer. Boostexter: A boosting-based system for text categorization. *Machine learning*, 39(2-3):135–168, 2000.
- [226] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [227] Yong-Soo Seol, Dong-Joo Kim, and Han-Woo Kim. Emotion recognition from text using knowledge-based artificial neural network. In *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications*, pages 1569–1572, 2008.
- [228] Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. A linguistic interpretation of the occ emotion model for affect sensing from text. In *Affective information processing*, pages 45–73. Springer, 2009.
- [229] Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita. Automatic summarization of twitter topics. In *National Workshop on Design and Analysis of Algorithms, Tezpur, India*, pages 121–128, 2010.
- [230] Kelly Shaver. *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media, 2012.
- [231] Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. A participant-based approach for event summarization using twitter streams. In *Proceedings of NAACL-HLT*, pages 1152–1162, 2013.
- [232] Liping Shen, Minjuan Wang, and Ruimin Shen. Affective e-learning: Using” emotional” data to improve learning in pervasive learning environment. *Educational Technology & Society*, 12(2):176–189, 2009.
- [233] Yang Shen, Shuchen Li, Ling Zheng, Xiaodong Ren, and Xiaolong Cheng. Emotion mining research on micro-blog. In *Web Society, 2009. SWS’09. 1st IEEE Symposium on*, pages 71–75. IEEE, 2009.
- [234] George Sher. *In praise of blame*. Oxford University Press, 2005.

- [235] Shiv Naresh Shivhare and Saritha Khethawat. Emotion detection from text. *arXiv preprint arXiv:1205.4944*, 2012.
- [236] Lidan Shou, Zhenhua Wang, Ke Chen, and Gang Chen. Sumblr: continuous summarization of evolving tweet streams. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 533–542. ACM, 2013.
- [237] Pranjal Singh and Divyanshu Bhartiya. Identification of certain emotions in text. 2012.
- [238] Push Singh et al. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, 2002.
- [239] Valentina Sintsovaa and Claudiu Musata. Fine-grained emotion recognition in olympic tweets based on human computation. *WASSA 2013*, 12, 2013.
- [240] John Jamieson Carswell Smart. Free-will, praise and blame. *Mind*, 70(279):291–306, 1961.
- [241] Koen Smets, Bart Goethals, and Brigitte Verdonk. Automatic vandalism detection in wikipedia: Towards a machine learning approach. In *AAAI workshop on Wikipedia and artificial intelligence: An Evolving Synergy*, pages 43–48, 2008.
- [242] Angela M Smith. Character, blameworthiness, and blame: comments on george shers in praise of blame. *Philosophical Studies*, 137(1): pages 31–39, 2008.
- [243] Angela M Smith. Moral blame and moral protest. *Blame: Its nature and norms*, Oxford University Press Oxford, pages 27–48, 2013.
- [244] Ahmad Soleimani and Ziad Kobti. Toward a fuzzy approach for emotion generation dynamics based on occ emotion model. *IAENG International Journal of Computer Science*, 41(1), 2014.
- [245] Efstathios Stamatatos. A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, 60(3): pages 538–556, 2009.

- [246] Bas R Steunebrink, Mehdi Dastani, and John-Jules Ch Meyer. The occ model revisited. In *Proceedings of the 4th Workshop on Emotion and Computing*, volume 65, pages 2047–2056, 2009.
- [247] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. *The general inquirer: A computer approach to content analysis*. MIT press 1966.
- [248] C. Strapparava and Valitutti A. Wordnet-affect: An affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086, 2004.
- [249] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 70–74, 2007.
- [250] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics, 2007.
- [251] Carlo Strapparava and Rada Mihalcea. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560. ACM, 2008.
- [252] Peter Frederick Strawson. *Freedom and resentment*, 1963.
- [253] Alistair Sutcliffe. User-oriented requirements engineering. In *Usability and Accessibility Focused Requirements Engineering (UsARE), 2014 IEEE 2nd International Workshop on*, pages 1–8. IEEE, 2014.
- [254] Maite Taboada. Discourse markers as signals (or not) of rhetorical relations. *Journal of pragmatics*, 38(4): pages 567–592, 2006.
- [255] Maite Taboada and William C Mann. Rhetorical structure theory: Looking back and moving ahead. *Discourse studies*, 8(3):423–459, 2006.
- [256] Matthew Talbert. Moral competence, moral blame, and protest. *The Journal of ethics*, 16(1):89–109, 2012.



- [257] Duyu Tang, Bing Qin, Ting Liu, and Qiuhui Shi. Emotion analysis platform on chinese microblog. *arXiv preprint arXiv:1403.7335*, 2014.
- [258] Yi-jie Tang and Hsin-Hsi Chen. Emotion modeling from writer/reader perspectives using a microblog dataset. In *Proceedings of IJCNLP Workshop on sentiment analysis where AI meets psychology*, pages 11–19, 2011.
- [259] Siddiqui Tanveer and U S Tiwary. *Natural language processing and Information Retrieval*. Oxford University Press, 2008.
- [260] David Martinus Johannes Tax. One-class classification. *PhD thesis, Delft University of Technology*, 2001.
- [261] Simo Editha Tchokni, Diarmuid O Séaghdha, and Daniele Quercia. Emoticons and phrases: Status symbols in social media. In *ICWSM*, pages 485-494, 2014.
- [262] Zhi Teng, Fuji Ren, and Shingo Kuroiwa. Retracted: recognition of emotion with svms. In *Computational Intelligence*, pages 701–710. Springer, 2006.
- [263] Robert A. Thamm. *The Classification of Emotions*. Springer US, Boston, MA, 2006.
- [264] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1): pages 163–173, 2012.
- [265] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12): pages 2544–2558, 2010.
- [266] Neal Tognazzini and D. Justin Coates. Blame. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2016 edition, 2016.

- [268] Kristina Toutanova and Christopher D Manning. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*, pages 63–70. Association for Computational Linguistics, 2000.
- [269] Oren Tsur, Dmitry Davidov, and Ari Rappoport. The international conference on weblogs and social media (icwsm)-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *The International Conference on Weblogs and Social Media (ICWSM)*, 2010.
- [270] M Teresa Turell. The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *International Journal of Speech, Language & the Law*, 17(2), 2010.
- [271] Mika Turkia. A computational model of affects. *arXiv preprint arXiv:0811.0123*, 2008.
- [272] Orizu Udochukwu and Yulan He. A rule-based approach to implicit emotion detection in text. In *International Conference on Applications of Natural Language to Information Systems*, pages 197–203. Springer, 2015.
- [273] Benjamin Van Durme and Ashwin Lall. Streaming pointwise mutual information. In *NIPS*, pages 1892–1900, 2009.
- [274] Jean-Christophe Verstraete. Subjective and objective modality: Interpersonal and ideational functions in the english modal auxiliary system. *Journal of pragmatics*, 33(10): pages 1505–1528, 2001.
- [275] Inaki San Vicente, Rodrigo Agerri, and German Rigau. Q-wordnet ppv: Simple, robust and (almost) unsupervised generation of polarity lexicons for multiple languages. *arXiv preprint arXiv:1702.01711*, 2017.
- [276] David Vilares, Carlos Gómez-Rodríguez, and Miguel A Alonso. Universal, unsupervised (rule-based), uncovered sentiment analysis. *Knowledge-Based Systems*, 118: 45–55, 2017.
- [277] BKH Vo and NIGEL Collier. Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, 4(1): pages 159–173, 2013.

- [278] R Jay Wallace. *Responsibility and the moral sentiments*. Harvard University Press, 1994.
- [279] R Jay Wallace. Dispassionate opprobrium: On blame and the reactive sentiments. *Reasons, recognition: Essays on the philosophy of TM Scanlon*, pages 384–385, 2011.
- [280] Joseph B Walther, Tracy Loh, and Laura Granka. Let me count the ways the interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of language and social psychology*, 24(1): pages 36–65, 2005.
- [281] William Yang Wang and Kathleen R McKeown. Got you!: automatic vandalism detection in wikipedia with web-based shallow syntactic-semantic modeling. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1146–1154. Association for Computational Linguistics, 2010.
- [282] James C Ward and Amy L Ostrom. The internet as information minefield: An analysis of the source and content of brand information yielded by net searches. *Journal of Business research*, 56(11):907–914, 2003.
- [283] Gary Watson. Two faces of responsibility. *Philosophical Topics*, 24(2):227–248, 1996.
- [284] Andrew G West, Sampath Kannan, and Insup Lee. Detecting wikipedia vandalism via spatio-temporal analysis of revision metadata? In *Proceedings of the Third European Workshop on System Security*, pages 22–28. ACM, 2010.
- [285] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.
- [286] Wikipedia. Wikipedia, the free encyclopedia, 2016. [Online; accessed 22-January-2016] <http://en.wikipedia.org>.
- [287] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.

- [288] Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. Emotion recognition from text using semantic labels and separable mixture models. *ACM Transactions on Asian Language Information Processing (TALIP)*, 5(2):165–183, 2006.
- [289] Ye Wu and Fuji Ren. Improving emotion recognition from text with fractionation training. In *Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on*, pages 1–7. IEEE, 2010.
- [290] Wei Xu, Ralph Grishman, Adam Meyers, and Alan Ritter. A preliminary study of tweet summarization using information extraction. *NAACL 2013*, page 20, 2013.
- [291] DUAN YaJuan, CHEN ZhuMin WEIF uRu, ZHOU Ming Heung, and Yeung SHUM. Twitter topic summarization by ranking tweets using social influence and content quality. 2012.
- [292] Changhua Yang, Kevin H Lin, and Hsin-Hsi Chen. Emotion classification using web blog corpora. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 275–278. IEEE, 2007.
- [293] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Writer meets reader: Emotion analysis of social media from both the writer’s and reader’s perspectives. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT ’09*, pages 287–290, Washington, DC, USA, 2009. IEEE Computer Society.
- [294] Hao Yaojun, Wang Jianguo, and Zhao Qingshan. Learning support technology: Design and implementation with the use of personality-occ model. *Distance Education in China*, 2:007, 2013.
- [295] Eunsang Yoon, Hugh J Guffey, and Valerie Kijewski. The effects of information and company reputation on intentions to buy a business service. *Journal of Business Research*, 27(3):215–228, 1993.
- [296] Honglei Zeng, Maher A Alhossaini, Li Ding, Richard Fikes, and Deborah L McGuinness. Computing trust from revision history. Technical report, Stanford Knowledge Systems LAB, 2006.

- [297] Jianwei Zhang, Yukiko Kawai, Shinsuke Nakajima, Yoshifumi Matsumoto, and Katsumi Tanaka. Sentiment bias detection in support of news credibility judgment. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pages 1–10. IEEE, 2011.
- [298] Xin Wayne Zhao, Yanwei Guo, Rui Yan, Yulan He, and Xiaoming Li. Timeline generation with social attention. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 1061–1064. ACM, 2013.