An Angle-based Interest Model for Text Recommendation

Bei Xu and Hai Zhuge

Institute of Cyber-Physical-Social Intelligence, Nanjing University of Posts and Telecommunications, China
Aston University, Birmingham, B4 7ET, UK
Key Lab of Intelligent Information Processing, Chinese Academy of Sciences, Beijing, China

Abstract

Building an interest model is the key to realize personalized text recommendation. Previous interest models neglect the fact that a user may have multiple angles of interests. Different angles of interest provide different requests and criteria for text recommendation. This paper proposes an interest model that consists of two kinds of angles: persistence and pattern, which can be combined to form complex angles. The model uses a new method to represent the long-term interest and the short-term interest, and distinguishes the interest on object and the interest on the link structure of objects. Experiments with news-scale text data show that the interest on object and the interest on link structure have real requirements, and it is effective to recommend texts according to the angles.

Keywords: Text recommendation; Interest model; Multi-Angle interest

1. Introduction

A key issue in personalized text recommendation is how to construct an appropriate user model to represent and adapt to user interests and then recommend texts according to the model. So a user model is often constructed by using personal reading history.

The keyword-based methods are the most commonly used methods to represent user interests [12, 13, 14, 19]. Keywords are extracted from text or provided by users, e.g., in iGoogle. Keywords are often calculated with weights [15], e.g. *tf-idf* [16, 20, 31]. Relevant texts are recommended through the techniques on words' weights, e.g. using cosine [16] or collaborative-filtering [3, 18, 21]. Keywords can also be organized into topics to select the texts close to the interested topics [7, 8]. There are other kinds of keyword methods. Histogram methods use histogram to analyze the statistics of keywords [20, 22]. The tag methods allow a user to assign a text with personalized tags or well-designed evaluating indicators to indicate the user's interest in the text [23, 33], and then use collaborative filtering to analyze user evaluations [39].

The background knowledge methods incorporate concept hierarchies, ontology or encyclopedic knowledge to analyze a user's interests and help users express requirements [29, 34, 35, 36]. For example, Google News gives structured categories and recommends texts based on the categories. The interested domains sometimes are fixed and sometimes evolve with interests [28].

The network methods represent text as one or several networks where concepts or basic text units (e.g., words, concepts, sentences, paragraphs or texts) are considered as weighted nodes and edges between nodes are established if some conditions are satisfied (e.g. co-occurring) [19, 24]. A node's weight often represents the degree of preference or significance.

A key limitation of the previous methods is that they essentially neglect the angle of user interests. A user can have multiple angles of interests, and different angles correspond to different sets of interested texts.

Some angles of interests have been addressed in previous works. Some text recommendation systems classify angles in term of domains, e.g., Yahoo news. However, using background methods, a user may receive a broad range of texts in the interested domains that they are not interested in [1]. Some systems incorporate long-term and short-term interests. The systems ask user to label words or concepts with "*long-term*" or "*short-term*", or require users to provide some long-term/short-term interests for extracting features, or consider stable interested domains as long-term interests and fast-changing interested domains as short-term interests [9, 10, 31]. However, users sometimes need more precise services than domain-oriented recommendation and often request the system to discover the long-term/short-term interests automatically. Some systems train long-term/short-term models based on the distribution of words [38]. However they do not consider the connections between keywords.

There are other angles that previous researches have neglected, for example: (1) a user may be interested in the texts focusing on a specific object (a person, a place, etc.). If a user is interested in an object, any text concerning the object meets the user's need no matter what event the objects are involved in. For example, a user who is interested in US president *Bush* is interested in the texts related to *Bush* no matter what event the text describes, and (2) a user may be interested in a link structure of objects that contains not only objects but also the links between objects. Different link structures of the same set of objects may indicate different events. For example, given two reading histories h_1 and h_2 of two users u_1 and u_2, some texts describe the tax policy of president *Bush* and some describe the activities of *Bin Laden* in *Sudan* in h_1 while the texts in h_2 only describe the 911 attack. Neither of the two parts in h_1 directly relate to 911-attack. *Bush* and *Bin Laden* are more closely linked in h_2 rather than in h_1. The interests reflected by h_1 and h_2 separately match two link structures illustrated in Fig. 1, where S_1 is the partial link structure of the objects in history h_1, and S_2 is the partial link structure of the objects in history h_2.

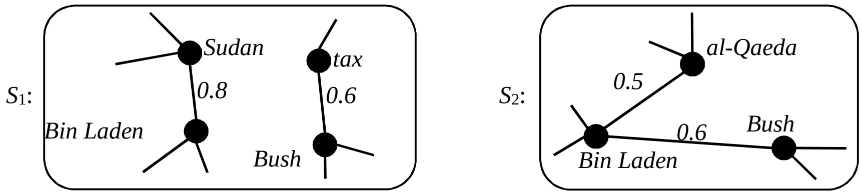


Fig. 1. Two partial link structures interested by two users, where a node denotes an object and an edge denotes that the two objects are linked to a certain degree.

The following is a short text related to 911-attack from *CNN*. It meets u_1's interest and u_2's interest on object because it shares a lot of objects with h_1 and h_2. But it does not meet u_1's interest on link structure because its structure and S_1 do not have any intersection of links even though they have many common objects, and it meets u_2's interest on link structure because its structure and S_2 have intersection.

WASHINGTON (CNN) -- Osama ***bin Laden*** *is the "prime suspect"in last Tuesday's terrorist attacks in New York and Washington and the United States wants to capture him, President* ***Bush*** *said Monday.*

Speaking with reporters after a Pentagon briefing on plans to call up reserve troops, ***Bush*** *offered some of his most blunt language to date when he was asked if he wanted* ***bin Laden*** *dead.*

*"I want justice,"***Bush*** *said. "And there's an old poster out West I recall, that said, 'Wanted, Dead or Alive.'"*

(from *http://edition.cnn.com/2001/US/09/17/bush.powell.terrorism/index.html?_s=PM:US*)

When a user wants to read the texts that contain a special link structure, previous methods probably recommend the texts that share a part of objects with the link structure. In fact, objects are mutually influenced through links. For example, if a user is interested in *George Bush* and already read several texts about Iraq war, then it is reasonable to presume that the user is interested in *Saddam*'s activities in Iraq war as well. This association influence should be considered in the interest model.

This paper proposes an angle-based interest model (*AIM*) with two kinds of angles and the complex angles combined by the angles for accurate text recommendation.

2. Related work

There are several relevant text recommendation techniques that build various kinds of user models.

The content-based recommendation methods can extract features from content and then recommend texts with similar features. The vector space model represents text as a vector and calculates the similarity between the vectors [20, 26, 31]. The components in a vector are words or phrases and the weights of the components usually are *tf-idf* values. The topic model analyses the term distributions (e.g., *PLSI* and *LDA*) and obtains embedded topics. There are some other content-based methods, e.g., Newsjunkie defines the information novelty to recommend texts with new stories [4].

The collaborative filtering methods can provide personalized service for a user based on the behaviors of similar users. Some methods recommend texts based on the rating of texts from other similar users [2, 9]. Some methods predicate user behaviors probabilistically based on the user's historical behaviors (e.g., click distribution) [6, 19, 32]. Collaborative filtering works well when the overlap between user behaviors is relatively high and the interests are relatively stable. So it is widely used in item recommendation on shopping website, like Amazon. Some researchers design hybrid methods to combine content-based recommendation and collaborative filtering [17, 30].

The context-aware text recommendation considers the context information [40], e.g., time or location [5]. Context information is usually attached to the interested concepts or interested topics. The context-aware text recommendation enriches the information around interests.

The above techniques do not differentiate multiple angles of interests, especially the interest on link structure. Another work similar to the angle-based interest is faceted navigation. Faceted navigation segments texts into pieces and organizes the pieces into facets [38, 42]. Each facet represents one aspect of the meaning of content. The pieces in a facet may come from different texts. Text recommendation based on angle interest and faceted navigation both classifies contents. The difference includes two aspects: (1) text recommendation based on angle interest concerns user while faceted navigation concerns text. (2) Faceted navigation can help narrow user interest while browsing, but it does not recommend the content of particular facets to particular user.

3. Interest model

Our interest model consists of the following angles of interest.

1. Pattern. Reading was regarded as a process of identifying objects and mapping them into a semantic image in mind [27]. One basic process of reading is to identify objects first and then establish the links between the objects. This is in line with the following phenomenon: a reader with a particular interested in some objects usually searching the objects in the text quickly without comprehensive reading, and a reader usually tends to read more closely if the reader is interested in a link structure (e.g., a specific event).

(1) **Object**. Object is the simplest pattern. An object maybe physical, e.g., a person, a place, or an abstract object. An object is normally indicated by a noun or noun phrase (no stop words) because they directly reflect objects while other types of words render objects. But not all nouns or noun phrases indicate objects. In real application, a list of non-meaningful nouns can help select the nouns that represent object. Wikipedia is useful to identify the names of objects, especially when the name is a noun phrase. We do not use WordNet because it is weak in identifying noun phrase. In the following, a word represents an object as well as meaningful noun (or noun phrase) if there is no special statement.

(2) **Link structure of objects**. A link structure represents a kind of co-occurrence of a set of objects and the links between objects. For example, if a reader is interested in a link "*Churchill* — *Hitler*", then a text that contains the common activities of *Churchill* and *Hitler* in *WWII* meets the interest, and texts that separately introduce the activities of *Churchill* and *Hitler* in *WWI* do not meet the interest. In searching, a user inputs two keywords, k_1 and k_2, to request the pages matching structure $k_1 - k_2$. If a user inputs three keywords, k_1, k_2 and k_3, then the input probably implies a structure $k_1 - k_2 - k_3 - k_1$.

2. Persistence. People's interests can be divided into two kinds according to the persistence: long-term interest and short-term interest. The characteristics of long-term and short-term interests are embodied in reading history. This paper considers a reading history as a sequence of texts that have been confirmed by reading behaviors like clicking hyperlinks.

If a user is interested in a pattern for a period of time, the user normally has two kinds of behaviors in the period: One is to read the texts related to the objects in the pattern, and the other is to be inclined to read the texts whose core meanings concern the pattern. So a common characteristic of the short-term interest and the long-term interest is that they satisfy the ***accumulation assumption*** — the degree of interest on a pattern increases with two factors: (1) the number of texts containing the pattern in reading history, (2) the rank of the pattern in each text. In spite of other factors, if object *a* and object *b* have the same times of appearance in a reading history, and *a* has higher rank than *b*, then *a* is more likely to be interested.

On the contrary, the words that appear in many texts of a reading history are very likely to indicate the interested pattern. There are three main kinds of the words: (1) stop word, which is ignored in the paper, (2) the words that indicate the interests, and (3) the words that are suitable in many situations because of the general meanings, e.g., "*people*", "*thing*". The third kind of words may disturb the discovery of the interests. Most words of the third kind are not recognized as objects according to the list of non-meaningful nouns. Ignoring the third kind of words does not influence discovering interest from reading history. One reason is that the number of the third kind of words is relatively small. Another key reason is that a user is normally not interested in a word of the third kind without being interested in the other words of the second kind. For example, if a user is interested in an airplane crush event, because most news on the event contains the sentences like "… people died …", then "*people*" appears many times in the reading history. But there must be other related words that can indicate the interest, such as "*airplane*", etc.

The key different characteristics of short-term interest and long-term interest are the distributions of the appearance of the interested pattern in reading history.

(1) **Short-term interest (*SI*)** indicates the inconstant and fast-changing interest. So the short-term interests are reflected by the patterns in the texts of recent reading history. The short-term interest satisfies a **proximity distribution assumption** — a pattern is very likely to be short-term interested if it appears in a condense way in the recent texts of the reading history, and the more a pattern is close to the current time, the more representative the pattern is for the short-term interest. A descending function is applied to describe the assumption.

(2) **Long-term interest (*LI*)** indicates the persistent interest that lasts for a relatively long time. The long-term interested patterns appear in a wide range of the reading history if the history lasts for a relatively long time. Every time a pattern appears in the reading history, it indicates that, in the time around the appearance, the texts concerning the pattern meet the user's interest to a certain degree. So, a user has a longer time being interested in a pattern if the pattern appears in two separated texts compared with appearing in two adjacent texts. Fixing the number of texts containing a pattern, the evener distribution the pattern has in the reading history, the longer time the pattern is being interested. So the long-term interest satisfies **wide distribution assumption** — the pattern that a person is long-term interested in is widely distributed in the reading history. Considering two patterns with the same times of appearances, the pattern that is evenly distributed in the whole range of the reading history should be greater long-term interested than the pattern that is concentrated within a small range of the reading history. A distributing coefficient is designed to support the assumption. The details of distributing coefficient are presented in appendix B.

There are two other characteristics of the long-term and the short-term interest: (1) One interest can be long-term and short-term at the same time. For example, if a person is long-term and short-term interested in an object, the object should not only appear in recent texts of the reading history, but also be widely distributed in the whole range of the reading history. (2) Long-term and short-term interests can be transformed into each other. A short-term interest transforms into long-term if it comes to the person's mind from time to time, and a long-term interest transforms into short-term if it is recalled and appeared in recent texts.

The above four kinds of interests can be combined into four basic angles of interest as follows: (1) *short-term object interest*, (2) *long-term object interest*, (3) *short-term link interest*, and (4) *long-term link interest*. Each basic angle establishes a criterion for selecting the recommended texts.

Based on the four basic angles, one complex angle can be described as a combination of the basic angles. For example, a user who adores *Justin Bieber* for a long time wants to know some affairs he or she does not know, a text meets the interest if (1) the text involves *Justin Bieber* that is a long-term interested object, and (2) the affairs are new for the system user, which means the link structure of *Justin Bieber* and other objects is neither short-term nor long-term interested. In this case, the user has a complex angle aiming at the texts that match the new link structures containing the long-term interested objects. Meeting the complex angle implies (1) meeting the long-term object interest, (2) not meeting the short-term link interest, and (3) not meeting the long-term link interest. So the complex angle is combined by the second, the third and the fourth basic angles.

4. Extracting Interest model from Texts

4.1. Methods

Our idea is to extract the data structure that represents the interest from the texts in reading history through a text scanning mechanism considering the following points:

1. *Simulating human reading features to emerge interest*

Link is a way to emerge interest in reading. On one hand, a link can increase the possibility of generating a new interest from an existing interest. For example, if a user is interested in US president *Bush* and has read some texts on *Bush* and *Bin Laden*, then the user is likely to be interested in *Bin Laden* as well. The text scanning mechanism can simulate the association process in human reading process. On the other hand, people generate impressions on the objects in text, and then knit the impressions together through association to understand the text [41]. Association can enhance the impressions of the linked objects. For example, if a user has read some texts on *Romeo* and *Juliet*, the impression on *Romeo* will be enhanced when *Juliet* appears, and vice versa.

The text scanning mechanism is introduced in appendix *A*.

2. *The data structure for calculating the angles of interest*

We use a weighted graph to record and calculate interests, where the nodes indicate the objects and the undirected edges indicate the associations. Each node or edge has two weights: the short-term weight representing the degree of short-term interest on the node or edge and the long-term weight representing the degree the long-term interest.

The interest graph is dynamically constructed with the growth of reading history. The graph is not large because the limited number of existing nouns or noun phrases (in English, there are about 2000 commonly-used nouns or noun phrases).

The scanning of each text in reading history enriches the interest graph as shown in Fig. 2. The objects and edges are extracted from each text and added to the graph if they have not appeared before. The calculation of the short-term weight considers accumulation assumption and proximity distribution assumption. The calculation of the long-term weight considers accumulation assumption and wide distribution assumption. After scanning one text, the previous weights will be updated. The update process takes the incoming text as input and needs not to re-calculate the previous texts in the reading history. Given an object (or edge) in the interest graph before scanning the incoming text, if the incoming text contains the object (or edge), the weights of the object (or edge) obtained from the incoming texts are accumulated onto the previous weight during the update process. The two weights of each node or each edge separately represent the degree of the short-term interest and the long-term interest on the node or the edge. Greater weight reflects stronger short-term interest or long-term interest.

The construction of the interest graph and the calculation of the weights are presented in appendix *B*.

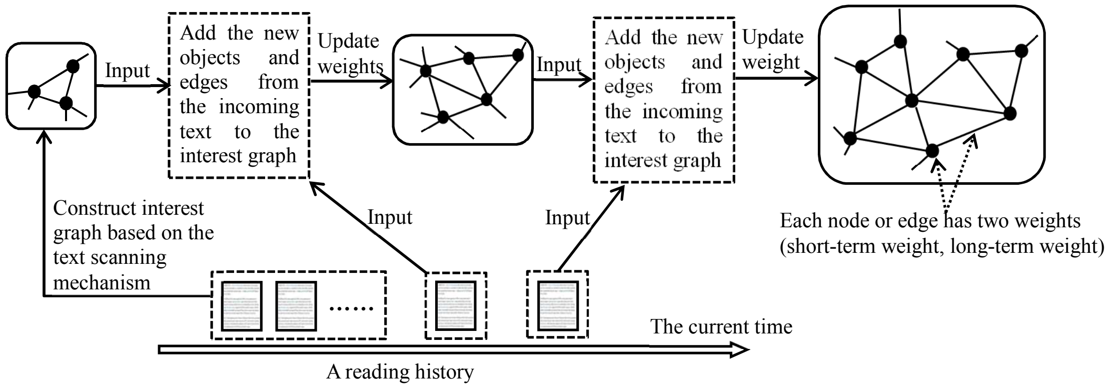

Input
Add the new objects and edges from the incoming text to the interest graph
Update weights
Input
Add the new objects and edges from the incoming text to the interest graph
Update weight
Construct interest graph based on the text scanning mechanism
Input
Input
Each node or edge has two weights (short-term weight, long-term weight)
......
The current time
A reading history

Fig. 2. The construction of interest graph. The sequence of the texts is a reading history. The hollow arrow denotes the direction of the

text scanning. The nodes and edges in interest graph are enriched with the scanning. The dotted arrows denote notes. The solid arrows denote the construction and the update of the interest graph. The dotted boxes denote the enriching processes

The nodes with relatively great short-term (or long-term) weights in the interest graph indicate the short-term (or long-term) interested objects. The degree of a text meeting the short-term (or long-term) object interest is determined by the intersection between the objects in the interest graph and the objects in the text as shown in Fig. 3. There are three attributes of the intersection: the size of the intersection, the short-term (or long-term) weights of the objects in the intersection, and the times of the appearance of the common objects in the text (one object may appear multiple times in the text).

A variable, *node match degree*, is used to measure the degree of a text meeting the short-term (or long-term) object interest. The calculation of node match degree considers the three attributes. If the short-term (or long-term) weight is considered as the second attribute, the node match degree measures the degree of the short-term (or long-term) object interest. So a text has two node match degrees with an interest graph. High node match degree means stronger interest on object.

The details of the node match degree are presented in appendix C.1.

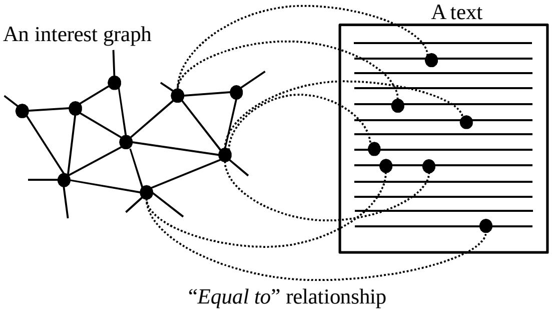


Fig. 3. The matching of the objects in a text and an interest graph. The circle nodes denote objects. The dotted curves denote "equal to" relationships and indicate the match between the objects in the interest graph and the words in the text. One object in the interest graph may appear many times in the text.

The interest on link structure is reflected by the nodes and the edges. The link connects two nodes to form a basic unit. A link structure consists of multiple basic units. The degree of the short-term (or long-term) link interest on a text is determined by the overlap between the basic units in the interest graph and the basic units in the text as shown in Fig. 4. A basic unit appears in a text if there is at least one sentence containing the two nodes. There are three attributes of the overlap. The first is the number of the basic units in the overlap. The second is the edges' short-term (long-term) weights of the basic units in the overlap. Because there may be multiple sentences in the text containing one basic unit, the third attribute is the times of the appearance of the common basic units in the text.

We use an *edge match degree* to measure the degree of a text meeting the short-term (or long-term) link interest. If the short-term (or long-term) weight is considered as the second attribute, the edge match degree measures the degree of short-term (or long-term) link interest. So a text has two edge match degrees with an interest graph. High edge match degree means stronger interest on link structure.

The details of edge match degree are presented in appendix C.2.

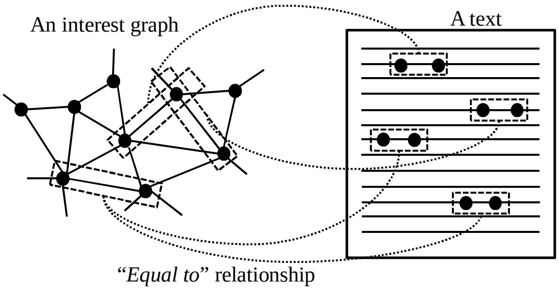


Fig. 4. The matching of the basic units in a text and an interest graph. Each horizontal line in the text represents one sentence. The circle nodes denote objects. The dotted boxes denote the basic units. The dotted curves denote "equal to" relationships and indicate the match between the basic units in the interest graph and the basic units in the text. One match unit in the interest graph may appear many times in the text.

3. *Recommendation mechanism*

The recommendation mechanism ranks the candidate texts by combining the four basic angles.

Each basic angle corresponds to a rank of texts. For a complex angle combined by *k* basic angles and the k basic angles correspond to *k* ranks, a text has a rank array (*RA*) that consists of *k* components corresponding to the *k* locations of the text in the *k* ranks, denoted as $\overrightarrow{RA} = (r_1, r_2 \ldots r_k)$. .

One rank array can be transformed into one value (Integrated rank, *Ir*) by multiplying a coefficient vector $\overrightarrow{CV}$ = (c_1, $c_2 \ldots c_k$) where c_1+c$_2$+…+c_k = 1 and c_1, $c_2 \ldots c_k$ are within [0, 1]. It is calculated by equation (1),

$$Ir = \overrightarrow{CV} \times \overrightarrow{RA} = \sum_{i=1}^{k} c_i * r_i \qquad (1).$$

Then the unread texts can be ranked based on the integrated rank. Notice that the integrated rank may not be integer. The default value of $\overrightarrow{CV}$ is (1/*k*, 1/*k*…1/*k*). The coefficients in coefficient vector can be adjusted by user for more precise service. The recommendation mechanism saves the used settings of the coefficient vectors for future use.

The following gives two typical complex angles that can be handled by the recommendation mechanism.

(1) The angle that concerns the new information on the short-term interested objects. For example, a user was interested in US economy history and read some texts on Franklin Roosevelt's economy policies. Then, the user is likely to generate interest in Roosevelt (object). Then the user tends to know more about Roosevelt's other aspects except economy like his activities in *WWII*. There are two aspects of meaning in the angle: a) "*new*" means the high-rank texts should not concern economy, which is handled by "*Low node match degree with long-term weight*", and b) "*short-term interested objects*" means the high-rank texts should concern Roosevelt, which is handled by "*High node match degree with short-term weight*". Therefore, the complex angle is combined by the short-term object interest and the long-term object interest.

(2) The angle that concerns the new link structure containing the long-term interested objects. For example, a user who has a long-term interest in *David Beckham* wants to know more information that the user does not know. Meeting the complex angle implies: a) meeting the long-term object interest, handled by "*high node match degree with long-term weight*", b) not meeting the short-term link interest, handled by "*low edge match degree with short-term weight*", and c) not meeting the long-term link interest, handled by "*low edge match degree with long-term weight*".

4.2. System Architecture

The system contains the following components: (1) an interface that monitors user behaviors, (2) a reading history record base, (3) a text scanning mechanism for scanning user's reading history, (4) an interest calculation mechanism that calculates the interest graph, (5) a recommendation mechanism that recommends texts from the external text repository, and (6) an external text repository that gathers texts from the Internet.

Fig. 5 shows the system architecture.

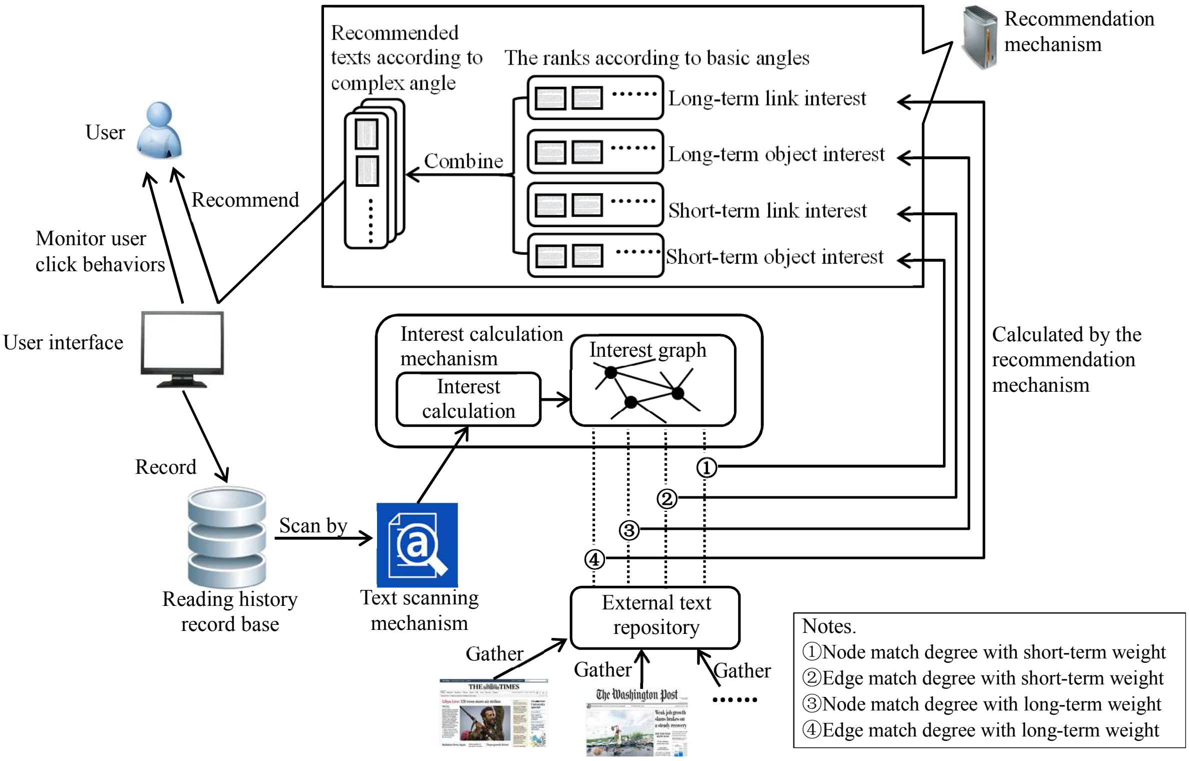

Recommended texts according to complex angle
The ranks according to basic angles
Recommendation mechanism
User
Combine
Long-term link interest
Long-term object interest
Short-term link interest
Short-term object interest
Recommend
Monitor user click behaviors
User interface
Interest calculation mechanism
Interest graph
Interest calculation
Calculated by the recommendation mechanism
Record
①
②
③
④
Scan by
Reading history record base
Text scanning mechanism
External text repository
Gather
Gather
Gather
······
Notes.
①Node match degree with short-term weight
②Edge match degree with short-term weight
③Node match degree with long-term weight
④Edge match degree with long-term weight

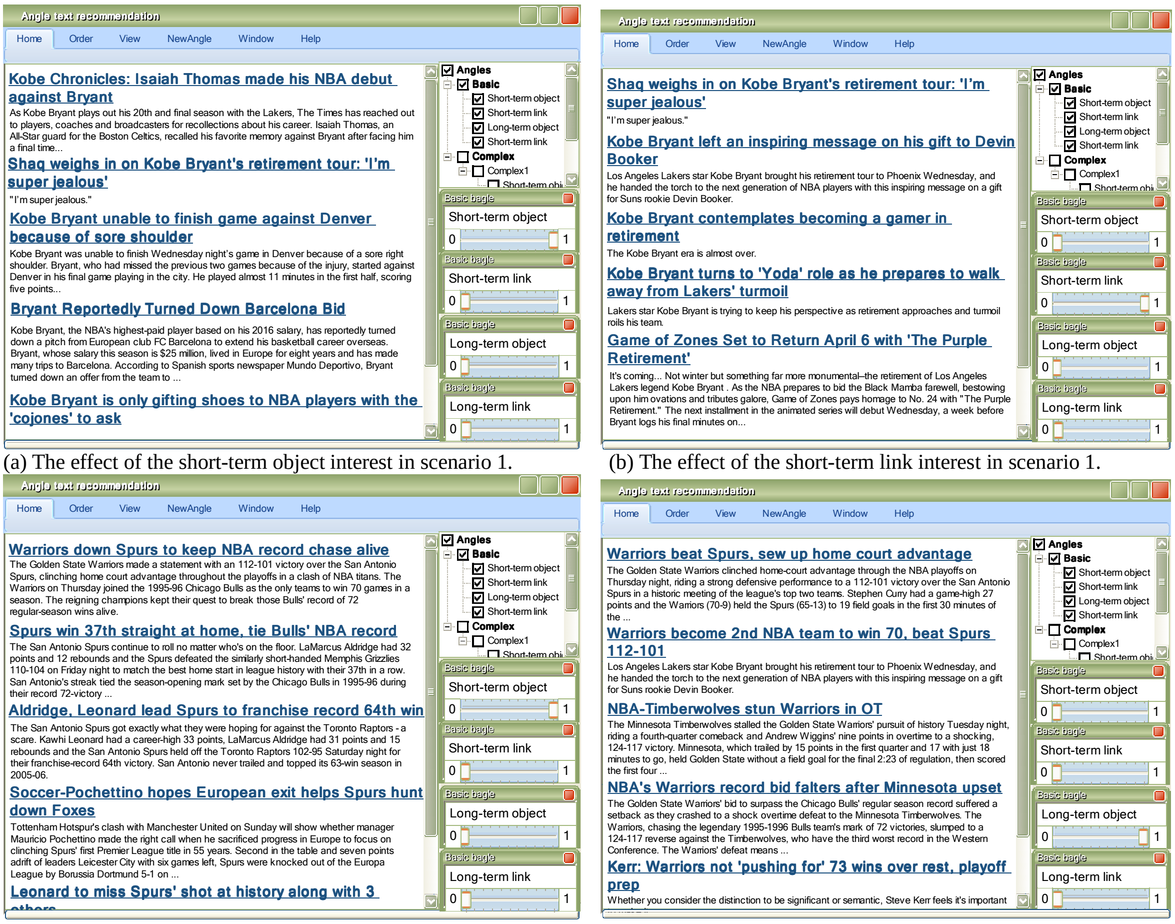
Fig. 5. The system architecture. The dotted lines denote the match between the candidate texts and the interest graph. The basic angles can combine into multiple complex angles based on different coefficient vectors.

The following are two typical application scenarios of the system.

Scenario 1. The effect of switching link interest to object interest. The user has recently read some news on the retirement of NBA player Kobe Bryant and sets the angle as the short-term object interest. The coefficient vector is (1, 0, 0, 0) as shown in the controllers in Fig. 6 (a). The recommended pieces of news concern other aspects of information on Kobe as well as the retirement event as shown in Fig. 6 (a). If the user wants to focus on the retirement event, he or she can adjust the angles to the short-term link interest by setting the coefficient vector as (0, 1, 0, 0), the recommended pieces of news only concern the retirement event as shown in Fig. 6 (b).

Scenarios 2. The effect of switching short-term interest to long-term interest. The user in this scenario often reads the news on NBA team Warriors in last year and read some news on team Spurs in recent days. If the user sets the coefficient vector as (1, 0, 0, 0), the recommended text are related to Spurs as shown in Fig. 6 (c). If the user wants to avoid the influence of recent reading behaviors and obtain information on the objects with long-term interest, he or she can adjust the coefficient vector to (0, 0, 1, 0). The recommended texts are shown in Fig. 6 (d).

Scenarios 3. A user does not know the interest angles, for example he/she does not know what is the short-term object interest and what is long-term object interest. The system can let the user know the angles he/she is potentially interested in by displaying the basic angles of interest. User can view, select and add his/her angles, and thereafter he/she can operate his/her potential angles.



☑ Angles
⊟ ☑ Basic
☑ Short-term object
☑ Short-term link
☑ Long-term object
☑ Short-term link
⊟ ☐ Complex
⊟ ☐ Complex1
☐ Short-term obj

Short-term object 0 1
Short-term link 0 1
Long-term object 0 1
Long-term link 0 1

Kobe Chronicles: Isaiah Thomas made his NBA debut against Bryant
As Kobe Bryant plays out his 20th and final season with the Lakers, The Times has reached out to players, coaches and broadcasters for recollections about his career. Isaiah Thomas, an All-Star guard for the Boston Celtics, recalled his favorite memory against Bryant after facing him a final time...

Shaq weighs in on Kobe Bryant's retirement tour: 'I'm super jealous'
"I'm super jealous."

Kobe Bryant unable to finish game against Denver because of sore shoulder
Kobe Bryant was unable to finish Wednesday night's game in Denver because of a sore right shoulder. Bryant, who had missed the previous two games because of the injury, started against Denver in his final game playing in the city. He played almost 11 minutes in the first half, scoring five points...

Bryant Reportedly Turned Down Barcelona Bid
Kobe Bryant, the NBA's highest-paid player based on his 2016 salary, has reportedly turned down a pitch from European club FC Barcelona to extend his basketball career overseas. Bryant, whose salary this season is $25 million, lived in Europe for eight years and has made many trips to Barcelona. According to Spanish sports newspaper Mundo Deportivo, Bryant turned down an offer from the team to ...

Kobe Bryant is only gifting shoes to NBA players with the 'cojones' to ask

(a) The effect of the short-term object interest in scenario 1.

Shaq weighs in on Kobe Bryant's retirement tour: 'I'm super jealous'
"I'm super jealous."

Kobe Bryant left an inspiring message on his gift to Devin Booker
Los Angeles Lakers star Kobe Bryant brought his retirement tour to Phoenix Wednesday, and he handed the torch to the next generation of NBA players with this inspiring message on a gift for Suns rookie Devin Booker.

Kobe Bryant contemplates becoming a gamer in retirement
The Kobe Bryant era is almost over.

Kobe Bryant turns to 'Yoda' role as he prepares to walk away from Lakers' turmoil
Lakers star Kobe Bryant is trying to keep his perspective as retirement approaches and turmoil roils his team.

Game of Zones Set to Return April 6 with 'The Purple Retirement'
It's coming... Not winter but something far more monumental–the retirement of Los Angeles Lakers legend Kobe Bryant . As the NBA prepares to bid the Black Mamba farewell, bestowing upon him ovations and tributes galore, Game of Zones pays homage to No. 24 with "The Purple Retirement." The next installment in the animated series will debut Wednesday, a week before Bryant logs his final minutes on...

(b) The effect of the short-term link interest in scenario 1.

Warriors down Spurs to keep NBA record chase alive
The Golden State Warriors made a statement with an 112-101 victory over the San Antonio Spurs, clinching home court advantage throughout the playoffs in a clash of NBA titans. The Warriors on Thursday joined the 1995-96 Chicago Bulls as the only teams to win 70 games in a season. The reigning champions kept their quest to break those Bulls' record of 72 regular-season wins alive.

Spurs win 37th straight at home, tie Bulls' NBA record
The San Antonio Spurs continue to roll no matter who's on the floor. LaMarcus Aldridge had 32 points and 12 rebounds and the Spurs defeated the similarly short-handed Memphis Grizzlies 110-104 on Friday night to match the best home start in league history with their 37th in a row. San Antonio's streak tied the season-opening mark set by the Chicago Bulls in 1995-96 during their record 72-victory ...

Aldridge, Leonard lead Spurs to franchise record 64th win
The San Antonio Spurs got exactly what they were hoping for against the Toronto Raptors - a scare. Kawhi Leonard had a career-high 33 points, LaMarcus Aldridge had 31 points and 15 rebounds and the San Antonio Spurs held off the Toronto Raptors 102-95 Saturday night for their franchise-record 64th victory. San Antonio never trailed and topped its 63-win season in 2005-06.

Soccer-Pochettino hopes European exit helps Spurs hunt down Foxes
Tottenham Hotspur's clash with Manchester United on Sunday will show whether manager Mauricio Pochettino made the right call when he sacrificed progress in Europe to focus on clinching Spurs' first Premier League title in 55 years. Second in the table and seven points adrift of leaders Leicester City with six games left, Spurs were knocked out of the Europa League by Borussia Dortmund 5-1 on ...

Leonard to miss Spurs' shot at history along with 3 others

(c) The effect of the short-term object interest in scenario 2.

Warriors beat Spurs, sew up home court advantage
The Golden State Warriors clinched home-court advantage through the NBA playoffs on Thursday night, riding a strong defensive performance to a 112-101 victory over the San Antonio Spurs in a historic meeting of the league's top two teams. Stephen Curry had a game-high 27 points and the Warriors (70-9) held the Spurs (65-13) to 19 field goals in the first 30 minutes of the ...

Warriors become 2nd NBA team to win 70, beat Spurs 112-101
Los Angeles Lakers star Kobe Bryant brought his retirement tour to Phoenix Wednesday, and he handed the torch to the next generation of NBA players with this inspiring message on a gift for Suns rookie Devin Booker.

NBA-Timberwolves stun Warriors in OT
The Minnesota Timberwolves stalled the Golden State Warriors' pursuit of history Tuesday night, riding a fourth-quarter comeback and Andrew Wiggins' nine points in overtime to a shocking, 124-117 victory. Minnesota, which trailed by 15 points in the first quarter and 17 with just 18 minutes to go, held Golden State without a field goal for the final 2:23 of regulation, then scored the first four ...

NBA's Warriors record bid falters after Minnesota upset
The Golden State Warriors' bid to surpass the Chicago Bulls' regular season record suffered a setback as they crashed to a shock overtime defeat to the Minnesota Timberwolves. The Warriors, chasing the legendary 1995-1996 Bulls team's mark of 72 victories, slumped to a 124-117 reverse against the Timberwolves, who have the third worst record in the Western Conference. The Warriors' defeat means ...

Kerr: Warriors not 'pushing for' 73 wins over rest, playoff prep
Whether you consider the distinction to be significant or semantic, Steve Kerr feels it's important

(d) The effect of the long-term object interest in scenario 2.

Fig. 6. The interface for switching angles.

5. Experiments

We apply the angle-based interest model to text recommendation to evaluate its effectiveness.

Experiment 1. The experiments on basic angles.
Purpose. This experiment is to demonstrate that the interest model can effectively recommend text according to the basic angles.
Dataset and preprocessing. The texts in real reading history normally come from various angles and the texts meeting one angle account for a small proportion of reading history. So using real user historic data has two limitations if one angle is tested: the real data may not reflect the angle we want to test and the texts recommended according to one angle have big difference with the real historic data. Therefore we manually design reading histories and appropriate texts for each angle. The texts come from the following two datasets.
(1) *Data-Test* 1. The data are the news from *DUC* 2005, 2006 and 2007. In the data, texts are divided into topics. The texts about a topic are selected from different newspapers or websites. We randomly choose 50 topics and 500 texts. Each topic contains 10 texts.
(2) *Data-Test* 2. The data are gathered from the Internet (mostly from the links in Wikipedia pages). The texts are related to *George W. Bush* and *Saddam*. *Data-Test* 2 contains 5 sets of texts that satisfy: (1) *Bush* and *Saddam* are not linked in the link structures of the texts in the first 4 sets. (2) *Bush* and *Saddam* are closely linked in the link structures of the texts in the fifth set.

Baselines. There are three baseline methods. The first is a content-based recommendation method (*CBM*) which uses keyword-based vector space model [31]. Based on *CBM*, the second baseline method (*Time-CBM*) takes time into consideration to make a targeted comparison by multiplying each word's weight with the descending function. The third baseline method (*Dis-CBM*) takes distribution into consideration by multiplying each word's weight with the word's distributing coefficient.
Procedure. For each angle, the experiment contains the following four steps: (1) set a reading history in which the texts reflect the angle, (2) set a group of ideal texts that meet the angle based on the reading history, (3) use the angle-based interest model to get the recommended texts, (4) compare the match ratios between the ideal texts and the recommended texts. Match ratio (*MR*) is designed as following, match ratio = *the number of the recommended texts in ideal result / the number of recommended texts*

Fig. 7 shows the results. More detailed experiments are presented in appendix D.

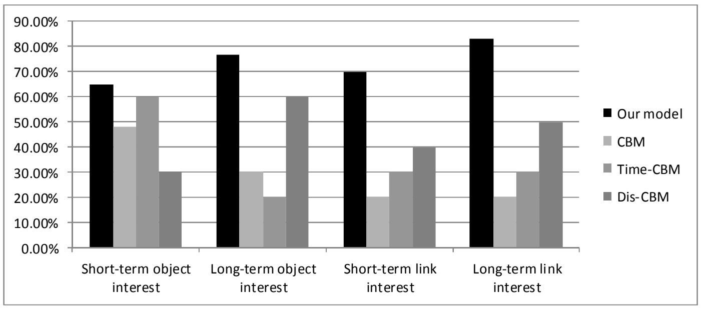


Fig. 7. The results of the experiment on basic angles. The vertical axis represents the match ratios. 10 texts are recommended.

The results show that the angle-based interest model outperforms the baselines. The advantage is not great when the recommended text according to object interest because the object angle is similar to the angle of previous text recommendation models. The advantage is great when the recommend text according to the link interest.

Experiment 2. The experiments on complex angles.
Purpose. The experiments is to test the effectiveness of the angle-based interest model handling complex angles (we select the angle that concerns the new information on the short-term interested objects and the angle that concerns the new link structure containing the long-term interested objects).
Dataset. *Data-Test* 1 and *Data-Test* 2.
***Procedure.* The following is the procedure:**
(1) Set a reading history in which the texts reflect the angle.
(2) Set a group of ideal texts that meet the angle based on the reading history.
(3) Set coefficient vector as default values.
(4) Use the angle-based interest model to get the recommended texts.
(5) Compare the ideal texts and the recommended texts.
(6) Change the coefficient vector and go to (4).

The details of the experiments are shown in appendix E. The results show that the angle-based interest model is effective to handle the two complex angles. The recommended texts can correctly change with the coefficient vector.

Experiment 3. The experiment on real world data
Purpose. The experiment demonstrates that the basic angles have real requirements and taking them into consideration can improve the effectiveness of text recommendation.
Dataset and preprocessing. Yahoo news dataset is used as real world data in which each record stores an anonymous user's click stream on the news. The data is preprocessed as following:
(1) Remove the click streams less than 50 news.
(2) Remove the click streams on which the recommendation effect is not improved while incorporating the short-term link interest and long-term link interest. The reason of the second preprocess is that not every user

have the interest on the link structure. Our interest model is useful as long as it can improve partial data. There are 373 click streams and 2270 different news after preprocess.

Baselines. We integrate the basic angles with two commonly used methods: content-based recommendation method (*CBM*) [31] and collaborative filtering method (*CFM*) [21]. The integrated methods are listed as following (*SO* stands for short-term object interest, *LO* stands for long-term object interest, *SP* stands for short-term link interest, *LP* stands for long-term link interest).

CBM - 1: 100% news from *CBM*.

(*CBM* + *AIM*) - 2: $1/\mu$ news from *SO* + $1/\mu$ news from *LO* + rest news from *CBM*.

(*CBM* + *AIM*) - 3: $1/\mu$ news from *SP* + $1/\mu$ news from *LP* + rest news from *CBM*.

(*CBM* + *AIM*) - 4: $1/2\mu$ news from *SO* + $1/2\mu$ news from *LO* + $1/2\mu$ news from *SP* + $1/2\mu$ news from *LP* + rest news from *CBM*.

CFM - 1: 100% news from *CFM*.

(*CFM* + *AIM*) - 2: $1/\mu$ news from *SO* + $1/\mu$ news from *LO* + rest news from *CFM*.

(*CFM* + *AIM*) - 3: $1/\mu$ news from *SP* + $1/\mu$ news from *LP* + rest news from *CFM*.

(*CFM* + *AIM*) - 4: $1/2\mu$ news from *SO* + $1/2\mu$ news from *LO* + $1/2\mu$ news from *SP* + $1/2\mu$ news from *LP* + rest news from *CFM*.

CBM - 1 and *CFM* - 1 are considered as baselines. (*CBM* + *AIM*) - 2 and (*CFM* + *AIM*) - 2 incorporate the interest on object. (*CBM* + *AIM*) - 3 and (*CFM* + *AIM*) -3 incorporate the interest on link structure. (*CBM* + *AIM*) - 4 and (*CFM* + *AIM*) - 4 incorporate the two angles.

Procedure. For one click stream, we consider the first 80% news as the reading history and the rest 20% news as the target news. The target news and 500 irrelevant texts form the unread text pool. For each method, the top 10, 20, and 30 news are recommended from the unread text pool. If the number of recommended news is not an integer, a rounding up value is taken.

The results are shown in Table 1.

Table 1. The experimental results on real data (P stands for precision, R stands for recall, F stands for F1-measure). $\mu = 6$.

Methods:	Top 10			Top 20			Top 30		
	P	R	F	P	R	F	P	R	F
CBM – 1:	0.327	0.156	0.211	0.272	0.259	0.265	0.221	0.315	0.26
(*CBM* +*AIM*) – 2:	0.325	0.154	0.209	0.281	0.257	0.268	0.244	0.337	0.283
(*CBM* +*AIM*) – 3:	0.382	0.178	0.243	0.302	0.268	0.284	0.249	0.351	0.291
(*CBM* +*AIM*) – 4:	0.372	0.171	0.234	0.309	0.294	0.301	0.277	0.396	0.326
CFM – 1:	0.338	0.161	0.218	0.277	0.232	0.253	0.223	0.319	0.262
(*CFM* +*AIM*) – 2:	0.341	0.162	0.22	0.283	0.264	0.27	0.26	0.371	0.306
(*CFM* +*AIM*) – 3:	0.366	0.172	0.234	0.291	0.277	0.284	0.258	0.369	0.304
(*CFM* +*AIM*) – 4:	0.358	0.17	0.231	0.298	0.284	0.291	0.269	0.384	0.316

Table 1 indicates the following points.

1. Because the integrated methods can improve the performance, so the interest on object and the interest on the link structure exist in users' needs, and the angles can be effectively handled by the angle-based interest model.
2. When the top 10 news are recommended, the advantages of (*CBM* + *AIM*) - 4 and (*CFM* + *AIM*) - 4 to the baselines are not great; but the advantages become greater when a bigger range is set (top 20, top 30). It indicates that, with the expanding of the range, incorporating the interest on object and the interest on link structure can provide more needed texts compared with *CBM* -1 and *CFM* - 1.
3. Compared with the baselines, (*CBM* + *AIM*) - 3 and (*CFM* + *AIM*) - 3 have greater advantages than (*CBM* + *AIM*) - 2 and (*CFM* + *AIM*) - 2 at the range of top 10. It indicates that incorporating the interest on link structure gives greater improvement than incorporating the interest on object while recommending a few texts. The improvement becomes weak with the expanding of the range. A reasonable explanation is that the interest on link structure appears less frequently than the interest on object.

Experiment 4. The experiment on the coefficients

Purpose. This experiment is to test the performance of the angle-based interest model when key coefficients change.

Dataset. *Data-Test* 1 and *Data-Test* 2.

Procedure. The experiments in experiment 1 are re-conducted by given different values of the coefficients. The details of the experiments are shown in Appendix F.

6. Conclusion

This paper proposes an interest model with two kinds of angles: persistence and pattern. It has the following characteristics:

1. It can distinguish the four basic angles of interest from a user's reading history.
2. The interest on object and interest on link structure are required by users. Incorporating the two angles is helpful to improve the effectiveness of text recommendation, and the interest on link structure appears less frequently than the interest on object.
3. It is better than the baseline methods while recommending texts according to the interest on object. The advantage is not great because the interest on object and the angle are similar to previous methods.
4. It is much better than the baseline methods while recommending texts according to the interest on link structure. The advantage is great especially while recommending a few texts. This is because previous methods do not consider the interest on link structure.
5. It can handle the complex angles. The effectiveness on the two typical complex angles (especially it is effective when the coefficient vector changes) strongly supports its effectiveness on the complex angles.

The matching according to the interest on link structure makes the content of recommended texts similar to the content of a reading history compared with the matching according to the interest on object. The matching can be closer, e.g., taking every three or more objects and their edges as one basic unit of link structure. But more similar matching does not always improve the effect of recommendation. Experiments show that the interest on link structure appears less frequently than the interest on object. If we use more similar matching, seldom texts can match with the reading history which makes the recommendation useless. On the other hand, text recommendation implies two characteristics: the recommended texts should be interested, and the information in the recommended texts should be more or less fresh. The two characteristics reflect two extremes. If only the first characteristic is emphasized, the recommendation is redundant, and if only the second characteristic is emphasized, the recommendation is irrelevant. More similar matching is inclined to emphasize the first characteristic.

Future researches include three aspects: (1) In addition to news-scale text, we will test the model on other types of text. (2) We will incorporate more angles into our model. (3) We will consider a semantics-rich pattern. (4) We will make use of the multi-dimensional resource space model to establish a multi-dimensional interest model [33]. The proposed model can also be extended to be a component of the semantic lens for personalization applications like summarization in the multi-dimensional cyber-physical society [11, 27, 37, 43, 44].

Acknowledgement

The research was supported by the National Science Foundation of China, the Science Foundation of Jiangsu (BK20140895) and the research funding from Nanjing University of Posts and Telecommunications (NY215119, NY213032, NY212003).

References

1. T. Lavie, M. Sela, I. Oppenheim, O. Inbar, and J. Meyer, "User Attitudes towards News Content Personalization," International Journal of Human-Computer Studies, vol. 68, pp. 483-495, 2010.
2. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," Proceedings of ACM Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.

3. J. Parra-Arnau, A. Perego, E. Ferrari, J. Forné, and D. Rebollo-Monedero, "Privacy-Preserving Enhanced Collaborative Tagging," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, 180–193, 2014.
4. E. Gabrilovich, S. Dumais, and E. Horvitz, "Newsjunkie: Providing personalized newsfeeds via analysis of information novelty," Proceedings of the 13th International Conference on World Wide Web, ACM, pp. 482–490, 2004.
5. W. Yao, J. He, G. Huang, J. Cao and Y. Zhang, "A Graph-based model for context-aware recommendation using implicit feedback data," World Wide Web journal, vol. 18, no. 5, pp. 1351-1371, 2015.
6. Ş. Gündüz and M. T. Özsu, "A web page prediction model based on click-stream tree representation of user behavior," Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 535-540, 2003.
7. W. Cui, S. Liu, L. Tan, et al., "Textflow: Towards better understanding of evolving topics in text," Visualization and Computer Graphics, IEEE Transactions on, vol. 17, no. 12, pp. 2412-2421, 2011.
8. M.P. Grineva, M.N. Grinev, and D. Lizorkin, "Extracting Key Terms from Noisy and Multitheme Documents," Proceedings of the 18th International Conference on World Wide Web, pp. 661-670, 2009.
9. B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," Proceedings of the 10th International Conference on World Wide Web, pp. 285–295, 2001.
10. K. Sugiyama, K. Hatano and M. Yoshikawa, "Adaptive Web Search based on User Profile Constructed without any Effort from Users," Proceedings 13th International Conference on World Wide Web, pp. 675-684, 2004.
11. H. Zhuge, The Knowledge Grid: Toward Cyber-Physical Society, World Scientific, 2nd Ed, 2012.
12. H. Sakagami and T. Kamba, "Learning Personal Preferences on Online Newspaper Articles from User Behaviors," Computer Networks and ISDN Systems, vol. 29, no. 8, pp. 1447-1455, 1997.
13. M. Balabanović and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Communications of the ACM, vol. 40, no. 3, pp. 66-72, 1997.
14. H. Lieberman, "Letizia: An Agent that Assists Web Browsing," Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI 95), pp. 924-929, 1995.
15. A. Moukas, "Amalthaea Information Discovery and Filtering Using a Multiagent Evolving Ecosystem," Applied Artificial Intelligence, vol. 11, no. 5, pp. 437-457, 1997.
16. G. Salton and M.J. McGill. Introduction to Modern Information Retrieval, New York, McGraw-Hill, 1983.
17. W. Chu and S. Park, "Personalized recommendation on dynamic content using predictive bilinear models," Proceedings of the 18th International Conference on World Wide Web, pp. 691–700, 2009.
18. G. Linden, B. Smith and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," Internet Computing, vol. 7, no. 1, pp. 76-80, 2003.
19. T. Hofmann, "Latent semantic models for collaborative filtering," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 89–115, 2004.
20. M.J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," The Adaptive Web, P. Brusilovsky, A. Kobsa, W. Nejdl, eds., Springer Berlin Heidelberg, pp. 325-341, 2007.
21. A.S. Das, M. Datar, A. Garg, and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering," Proceedings of the 16th International Conference on World Wide Web, pp.271-280, 2007.
22. M. Fredrikson and B. Livshits, "RePriv: Re-Envisioning In-Browser Privacy," Microsoft Research Technical Report, MSR-TR-2010-116, 2010.
23. J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, J.L. Muñoz, and O. Esparza, "Optimal Tag Suppression for Privacy Protection in the Semantic Web," Data & Knowledge Engineering, vol. 81-82, pp. 46-66, 2012.
24. F.A. Asnicar and C. Tasso, "ifWeb: a Prototype of User Model-Based Intelligent Agent for Document Filtering and Navigation in the World Wide Web," Proceedings of the 6th International Conference on User Modeling, pp. 3-11, 1997.
25. J. Chen, H. Zhuge, "Aimed information quantity in text," Concurrency and Computation: Practice and Experience, vol. 27, no. 15, pp. 3982-4000, 2014.
26. I. Soboroff and C. Nicholas, "Collaborative filtering and the generalized vector space model", Proceedings of the 23rd annual international ACM SIGIR conference on Research and developm, pp. 351-353, 2000.
27. H. Zhuge, "Interactive Semantics," Artificial Intelligence, vol. 174, pp. 190-204, 2010.
28. C.C. Chen, M.C. Chen, and Y. Sun, "PVA: A Self-Adaptive Personal View Agent," Journal of Intelligent Information Systems, vol. 18, no. 2-3, pp. 173-194, 2002.
29. M. Speretta and S. Gauch, "Personalized search based on user search histories," Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 622-628, 2005.
30. R. Burke, "Hybrid systems for personalized recommendations," Intelligent Techniques for Web Personalization, pp. 133–152, 2005.

31. P. Lops, M. De Gemmis, and G. Semeraro, "Content-Based Recommender Systems: State of the Art and Trends," Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, eds., 73-105, 2011.
32. D. Pennock, E. Horvitz, S. Lawrence and C. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory and model-based approach," Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, pp. 473–480, 2000.
33. H. Zhuge and Y. Xing, "Probabilistic Resource Space Model for Managing Resources in Cyber-Physical Society," IEEE Transactions on Service Computing, vol. 5, no. 3, pp. 404-421, 2012.
34. S.E. Middleton, N.R. Shadbolt, and D.C. De Roure, "Ontological User Profiling in Recommender Systems," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 54-88, 2004.
35. A. Csomai and R. Mihalcea, "Linking Documents to Encyclopedic Knowledge," IEEE Intelligent Systems, vol. 23, no. 5, pp. 34–41, 2008.
36. R. Mihalcea and A. Csomai, "Wikify!: Linking Documents to Encyclopedic Knowledge," Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management (CIKM), pp. 233-243, 2007.
37. H. Zhuge, "Semantic Linking through Spaces for Cyber-Physical-Socio Intelligence: A Methodology," Artificial Intelligence, vol. 175, pp. 988-1019, 2011.
38. D. H. Widyantoro, T. R. Ioerger and J. Yen. Learning user interest dynamics with a three-descriptor representation. Journal of the American Society for Information Science and Technology, vol. 52, no. 3, pp. 212-225, 2001.
39. B. Sarwar, G. Karypis, J. Konstan, & J. Riedl. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web, ACM, pp. 285-295, 2001.
40. J. Jancsary, F. Neubarth, and H. Trost, "Towards context-aware personalization and a broad perspective on the semantics of news articles," Proceedings of the 4th ACM conference on Recommender Systems, pp. 289–292, 2010.
41. B. Xu and H. Zhuge, "A Text Scanning Mechanism Simulating Human Reading Process," Proceedings of the 23th International Joint Conference on Artificial Intelligence (IJCAI 2013), pp. 2190-2196, 2013.
42. B. Xu and H. Zhuge, "Faceted Navigation through Keyword Interaction," World Wide Web Journal, vol. 17, pp. 671–689, 2014.
43. H. Zhuge, Multi-Dimensional Summarization in Cyber-Physical Society, Elsevier, 2016.
44. H. Zhuge, Cyber-Physical Society - The science and engineering for future society. Future Generation Computer Systems. Vol. 32, pp. 180-186, 2014.

Biography

Bei Xu is a lecturer of the Institute of Cyber-Physical-Social Intelligence at Nanjing University of Posts and Telecommunications, China. He was a PhD student of the Knowledge Grid research group at the Institute of Computing Technology, Chinese Academy of Sciences. His research interests concern text information processing and semantic link network. He has published a number of papers in international journals and conferences.

Hai Zhuge is a professor of Aston University in the UK, a joint professor of the Key Lab of Intelligent Information Processing in Chinese Academy of Sciences, and a guest professor of Nanjing University of Posts and Telecommunications in China. He has made systematic contribution to semantics modelling and knowledge management environment through lasting fundamental research on the semantic link network, the multi-dimensional category space and knowledge modelling. He is extending research towards Cyber-Physical Society, which concerns multi-disciplinary methodological, theoretical and technical innovation. Professor Zhuge is a Distinguished Scientist of the ACM (Association of Computer Machinery) for "Significant contribution and impact in computing field", a Distinguished Speaker of the ACM, a distinguished visiting fellow of Royal Academy of Engineering, and a Fellow of British Computer Society. He is serving as an associate editor of IEEE Intelligent Systems. Homepage: http://www.knowledgegrid.net/~h.zhuge.

Appendix A. Text Scanning Mechanism

The text scanning mechanism (*HTSM*) scans one text sentence by sentence and simulates the impression propagation process in association as shown in Fig. 8.

While scanning the k^{th} sentence,

1. Two ranges are set: the global range which is from the first sentence to the k^{th} sentence and the local range which is around the k^{th} sentence ($k \pm D$, D is the local size).

2. There are two kinds of relevancies between the words. (1) The global relevancy between two words w_1 and w_2, denoted as $GR_{\text{k}}(w_1, w_2)$, represents the relevancy between the two words in the global range while scanning the k^{th} sentence. It is measured by the number of sentences containing w_1 and w_2 in the global range. (2) The local relevancy between two words w_1 and w_2, denoted as $LR_{\text{k,D}}(w_1, w_2)$, represents the relevancy between the two words in the local range while scanning the k^{th} sentence. It is calculated by equation (2) and (3),

$$LR_{k,D}(w_1, w_2) = Min(N_{k,D}, 1) \times GR_k(w_1, w_2) \qquad (2),$$

$$N_{k,D} = \begin{cases} \dfrac{NumS_{k-D,k+D}(w_1, w_2)}{\log_r(NumS_{1,k}(w_1, w_2))} & \text{if}(k \in [D+1, N\text{-}D]) \\ 1 & \text{if}(k \in [1, D]) \\ \dfrac{NumS_{k-D,N}(w_1, w_2)}{\log_r(NumS_{1,N}(w_1, w_2))} & \text{if}(k \in [N\text{-}D+1, N]) \end{cases} \qquad (3),$$

where $N_{\text{k,D}}$ denotes the ratio of local relevancy and global relevancy; $NumS_{\text{x,y}}(w_1,w_2)$ denotes the number of sentences containing w_1 and w_2 from the x^{th} sentence to the y^{th} sentence; *Min*() is a function that returns the smaller value; and r is an integer. The values of the coefficients in equation (2) and (3) are discussed in [41].

3. A local word network (LWN_{k}) is built. The nodes are the words in the local range. The edges are built if the two words appear in a common sentence within the local range. An edge's weight is its local relevancy. A node's weight represents local word impression (*LWI*) of the word.

Association process is simulated in the network. The amount of propagated impression from node *a* to node *b* is determined by two factors: (1) the amount of impression that *a* passes out, and (2) the ratio between the local relevancy between *a* and *b* and the local relevancy between *a* and other neighbors.

The association process is performed as following. While scanning k^{th} sentence, for a word *w* in k^{th} sentence, *HTSM* adds a weight to *w*'s local word impression; *w* reserves a partial weight and propagates the rest to the neighbor nodes; the neighbors that receive weight do the propagation as well. The weight propagated through nodes is calculated by equation (4),

$$PW(i \to j) = \begin{cases} (1 - 1/\omega) \times W[i] \times \dfrac{LR_{k,D}(i, j)}{\sum_j LR_{k,D}(i, j)} & if(W[i] > MIN) \\ 0 \quad if(W[i] \le MIN) \end{cases} \qquad (4).$$

In equation (4), *j* denotes a neighbor node of *i*; *PW*(*i*→*j*) denotes the propagated weight from *i* to *j*; $LR_{\text{k,D}}(i, j)$ denotes the local relevancy between *i* and *j* after scanning the k^{th} sentence. *MIN* is a threshold that the propagated weight lower than *MIN* will be ignored. ω is a ratio that a word keeps $1/\omega$ weight and propagates $(1\text{-}1/\omega)$ weight to its neighbors. ω is set as 2.

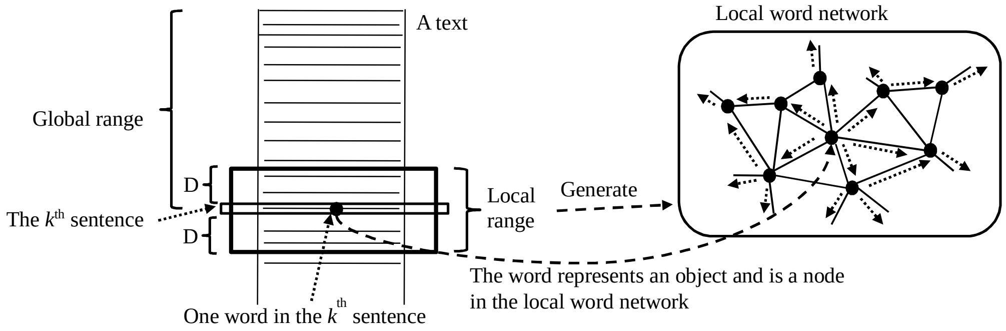

A text
Local word network
Global range
D
The kth sentence
D
Local range
Generate
The word represents an object and is a node in the local word network
One word in the kth sentence

Fig. 8. The simulation of association process. The dotted arrows in local word network represent the propagated impression. Each line in the text represents one sentence.

Appendix B. Data Structures for Extracting Interest from Texts

An interest graph consists of a set of nodes that are the objects appeared in the reading history and a set of edges between the nodes. If there is impression propagated between two nodes, the graph contains an edge between the two nodes.

The calculation of short-term weight considers accumulation assumption and proximity distribution assumption.

(1) Accumulation assumption requires to the ranks of the objects and the links in one text. If relatively great amount of impression is propagated through an object or a link, the object or the link is significant for understanding the meaning of the text and thusly has high rank. We give two concepts as following to measure the rank of object or link. One is impression throughput amount (*ITA*). Impression throughput amount of an object in a text is the sum of impression propagated through the corresponding word *w* after scanning the text. It is calculated by equation (5),

$$ITA_T(w) = \sum_i \sum_j PW_{T,i}(w \to w_j) \quad (5),$$

where *T* denotes a text, *w* and w_j denote two words in *T*, w_j is the j^{th} neighbor word that receives impression from *w*, $ITA_T(w)$ denotes the *ITA* of *w* in *T*, $PW_{T,i}(w \to w_j)$ denotes the weight propagated from *w* to w_j while scanning i^{th} sentence in *T* by using *HTSM.*

The other concept is impression flow amount (*IFA*). Impression flow amount of a link in a text is the sum of impression propagated between the corresponding two words after scanning the text. It is calculated by equation (6),

$$IFA_T(w_1, w_2) = \sum_j PW_{T,j}(w_1 \to w_2) + \sum_j PW_{T,j}(w_2 \to w_1) \quad (6),$$

where w_1 and w_2 denote two words in *T*, *T* denotes a text, $IFA_T(w_1,w_2)$ denotes the *IFA* between w_1 and w_2 in *T*, $PW_{T,j}(w_1 \to w_2)$ denotes the weight propagated from w_1 to w_2 while scanning j^{th} sentence in *T* by using *HTSM.*

If a user decides to read a text, the objects or the links with high rank in the text are very likely to be the interested things. However the objects and the links with high ranks do not always reflect a user’s interests. For example, a user reads a text describing 911-attack and the user is only interested in the actions of New York fire brigade, then the information on the terrorists does not meet the user’s interest even though the information has high rank in the text. So the *ITA* or *IFA* in a single text cannot be directly used to measure the degree of interest. The problem can be handled by considering a reading history instead of a single text. Even though an interested pattern does not have high rank in one text, it must be distributed in other texts in the reading history. So the *ITA* (or *IFA*) of an object (or a link) in a reading history are extended from a single value to an *ITA* array (or *IFA* array). The components in the array are the *ITA*/*IFA* values of the object/link in each text of the reading history. Then the proximity distribution assumption and wide distribution assumption can be applied on the arrays.

(2) A descending function is applied to implement proximity distribution assumption. The nodes appear in recent texts are assigned with relatively high weights. Other typical functions are discussed in the experiments.

According to accumulation assumption and proximity distribution assumption, the *ITA*s (or *IFA*s) of an object (or link) are summed up following a descending trend started from the current time. Given a reading history: T_1, T_2 … T_{SIZE}, where T_1 is the text most close to the current time, give two linked nodes n_1 and n_2, the short-term weight of the edge between n_1 and n_2 is calculated by equation (7) and (8),

$$SWE\left(n_1,n_2\right)=\sum_{k=1}^{SIZE}\Omega(k)*IFA_{T_k}\left(n_1,n_2\right) \tag{7},$$

$$\Omega\left(k\right) \; = \; 2/(1+\lambda^{k-1}) \qquad \lambda>1 \tag{8},$$

where k denotes the k^{th} text in the reading history. $SWE(n_1,\ n_2)$ denotes the short-term weight of the edge between n_1 and n_2 in the interest graph. $\Omega(k)$ is within [0, 1]. $\Omega(k)$ is a decreasing function that the decreasing speed of $\Omega(k)$ becomes slow when time lasts. The value of λ is set as 1.15. The values of λ are discussed. And $\Omega(k)$ is compared with other typical functions in the experiments. Similarly, the short-term weight of a node p is calculated by equation (9),

$$SWV(p)=\sum_{k=1}^{SIZE}\Omega(k)*ITA_{T_k}(p) \tag{9},$$

where $SWV(p)$ denotes the short-term weight of p in the interest graph.

The calculation of the long-term weight considers accumulation assumption and wide distribution assumption. Here, the accumulation assumption is the same with the accumulation assumption in the calculation of the short-term weight.

According to wide distribution assumption, a variable (distributing coefficient, *DC*) is designed to measure the degree of a word being evenly distributed in a reading history. The calculation of distributing coefficient takes an *ITA* (or *IFA*) array as input. For an array that consists of positive value or 0, distributing coefficient measures the degree of the positive values being evenly distributed. Two factors of *ITA*/*IFA* array are considered: (1) the consecutive zero sequences, and (2) the consecutive non-zero sequences. For example, in an array (2, 4, 0, 0, 5, 3, 2, 0, 0, 0, 3, 4), the three sequences containing the underlined values are two consecutive non-zero sequences and the rest two sequences are two consecutive zero sequences. The positive values are evenly distributed if the number of the consecutive zero and non-zero sequences is relatively big and the lengths of the sequences are relatively small. Two concepts are given to describe the two kinds of sequences. One is gap array (*GA*). For an array Q that consists of positive values and 0, the gap array of Q is an array (g_1, g_2, g_3 … g_{m}) where g_{i} denotes the number of 0 in the i^{th} consecutive zero sequence in Q. $GA(Q,\ i)$ denotes the i^{th} element in the gap array of Q. The other concept is inverse gap array ($\sim GA$). For an array Q that consists of positive values and 0, the inverse gap array of Q is an array (g_1, g_2, g_3, …, g_{m}) where g_{i} denotes the number of non-zero values in the i^{th} consecutive non-zero sequence in Q. $\sim GA(Q,\ i)$ denotes the i^{th} element in the inverse gap array of Q. For example, given an array Q = (85, 42, 0, 0, 0, 0, 33, 0, 101, 0, 0), the gap array is (4, 1, 2) and the inverse gap array is (2, 1, 1). $GA(Q,\ 2) = 1$ and $\sim GA(Q,\ 1) = 2$. If all values in Q are not 0, the gap array is (0); if all values in Q are 0, the inverse gap array is (0). Then the distributing coefficient of an array Q is calculated by equation (10),

$$DC(Q)=\log_{\alpha}((\sum_i(GA(Q,i))^m+\sum_j(\sim GA(Q,\mathrm{j}))^n)/N^m) \tag{10},$$

where m>n>1, 0<α<1, N is the number of components in Q. α is a coefficient. The default values of m, n and α are 3, 2 and 0.6. The values of m, n and α are discussed in the experiments. Because the distributing coefficient is to depict the degree of positive values being evenly distributed, gap array play more important role in the calculation than inverse gap array which leads to $m>n$. The distributing coefficient growths if the number of the sequences increases or the length of the sequences decreases.

Based on the distributing coefficient, two words' degrees of being evenly distributed can be compared as following. Given an *ITA* (or *IFA*) array A_p of node (or edge) p and an *ITA* (or *IFA*) array A_q of node (or edge) q, if $DC(A_p)$ is greater than $DC(A_q)$, the distribution of p is evener than the distribution of q. For example, given two nodes p and q, p's *ITA* array is A_p: (85, 42, 40, 0, 0, 0, 0, 0, 0, 20, 33, 0, 101, 0), node q's *ITA* array is A_q: (0, 42, 40, 0, 0, 85, 0, 0, 0, 20, 33, 0, 101, 0) which is obtained by exchanging the first and the 6th element in A_p. $DC(A_p) \approx 4.84$ and $DC(A_q) \approx 7.92$. So the distribution of q is evener than the distribution of p.

Based on the distribution coefficient and accumulation assumption, the long-term weights of the nodes and the edges are calculated as following. Given a reading history — $T_1, T_2 \ldots T_{SIZE}$, where T_1 is most close to the current time, let *Array_ITA*(p) be node p's *ITA* array, $Array\_ITA(p) = (ITA_{T_1}(p), ITA_{T_2}(p), \ldots, ITA_{T_{SIZE}}(p))$, and *DC*(*Array_ITA*($p$)) be the distributing coefficient of *Array_ITA*(p), the long-term weight of p is calculated by equation (11),

$$LWV(p) = DC(Array\_ITA(p)) \times \sum_{k=1}^{SIZE} ITA_{T_k}(p) \qquad (11),$$

where *DC*(*Array_ITA*(p)) reflects wide distribution assumption and $\Sigma\, ITA_{T_k}(p)$ reflects accumulation assumption. Similarly, the long-term weight of edge $e = (n_1, n_2)$ is calculated by equation (12),

$$LWV(n_1, n_2) = DC(Array\_IFA(n_1, n_2)) \times \sum_{k=1}^{SIZE} IFA_{T_k}(n_1, n_2) \qquad (12).$$

Appendix C. Match Degrees with Interest Model

C.1 Node match degree

Assuming L is an interest graph and t is a text, Z is the intersection between the objects in L and the objects in t, the match degree according to the interest on object is calculated by equation (13),

$$NMD_{t,L} = \frac{1}{N} \sum_{i=1}^{num} (T(n_i) \times wv_L(n_i)) \qquad (13),$$

where $NMD_{t,L}$ denotes the node match degree between t and L; N denotes the number of sentences in t; n_i denotes the ith element in Z; $T(n_i)$ denotes the frequency of n_i in t; if $NMD_{t,L}$ measures the degree of the short-term object interest, then $wv_L(n_i)$ denotes the short-term weight of n_i in L; if $NMD_{t,L}$ measures the degree of the long-term object interest, then $wv_L(n_i)$ denotes the long-term weight of n_i in L. The time complexity of the calculation is $O(M*N)$. M is the number of nodes in L. N is the number of object in t. It is a polynomial time complexity.

In equation (13), N is divided to avoid that a text has high node match degree just because it is long enough to cover many words.

C.2 Edge match degree

Assuming L is an interest graph, t is a text, and I is the intersection between the objects in L and the objects in t, the match degree according to the interest on link structure is calculated by equation (14),

$$EMD_{t,L} = \frac{1}{N} \sum_{i}^{num-1} \sum_{j=i+1}^{num} (T(n_i, n_j) \times we_L(n_i, n_j)) \qquad (14),$$

where $EMD_{t,L}$ is the edge match degree between t and L; n_i and n_j denote the ith and jth object in I; if $EMD_{t,L}$ measures the degree of short-term link interest, then $we_L(n_1, n_2)$ denotes the short-term weight of (n_1, n_2) in L; if $EMD_{t,L}$ measures the degree of long-term link interest, $we_L(n_1, n_2)$ denotes the long-term weight of (n_1, n_2) in L; if n_i and n_j are not linked in L, $we_L(n_1, n_2) = 0$; N denotes the number of sentences in t; $T(n_i, n_j)$ denotes the number of sentences in t containing n_i and n_j; *num* denotes the number of elements in I. The time complexity of the calculation is M*N+num^2+num^2*S = O(M*N*S). M is the number of nodes in L. N is the number of object in t. S is the number of sentences in t. It is a polynomial time complexity.

In equation (14), N is divided to avoid that a text has high edge match degree just because it is long enough to cover many basic units.

Appendix D. Experiments on Basic Angles.

D.1 Experiments on the interest on object

D.1.1 Experiment on short-term object interest

This experiment is to demonstrate that the angle-based interest model can recommend texts according to the node match degree with short-term weight.

From *Data-Test* 1, 20 topics are chosen and grouped into four sets — set_1, set_2, set_3 and set_4. Each set has 5 topics. 4 texts are randomly chosen from each topic. So 20 texts are chosen from one set and can be considered as a partial reading history with a random order. Four partial reading histories are obtained from the four sets and further merge into a whole reading history one after another as: $set^{20}_{1}\rightarrow set^{20}_{2}\rightarrow set^{20}_{3}\rightarrow set^{20}_{4}$ (total 80 texts in the whole reading history). set^{m}_{k} (k=1, 2, 3, 4) denotes the k^{th} partial reading history that contains m texts. In the whole reading history, the texts in set^{20}_{4} are most recently read; the texts in set^{20}_{3} are read before set^{20}_{4}, and so on. An interest graph can be obtained by scanning the whole reading history.

In the chosen 20 topics, 2 texts are chosen from the rest texts of each topic (total 40 texts). In the 40 texts, the texts from set_4 should be most relevant to the reading history according to the short-term object interest; the texts from set_3 should be second most relevant, and so on. In the other 30 topics, 3 texts are chosen from each topic (total 90 texts). The 90 texts are not relevant to the reading history. The 130 (40+90) texts form a testing set. The node match degree with short-term weight between each text in the testing set and the interest graph is calculated. The texts with the first i node match degrees are recommended (i = 10, 20, 30, 40 or 50).

To test the stability, the experiment is performed three times with different orders of the texts in the partial reading histories — denoted as $Order_1$, $Order_2$ and $Order_3$. The order of the partial reading histories does not change. Table 2 shows the experimental results where one value means the number of texts in the intersection between the texts with the first i node match degrees and the texts in the left set of the row. The values in the row "*other*" mean the number of the recommended texts not in set_1, set_2, set_3 or set_4.

The ideal results are: when i =10, there are 10 recommended texts from set_4, when i = 20, there are 10 recommended texts from set_4 and 10 recommended texts from set_3, and so on.

Table 2. The experimental results on the node match degree with short-term weight.

$Order_1$	i=10	i=20	i=30	i=40	i=50
set_4	9	10	10	10	10
set_3	1	5	8	10	10
set_2	0	4	7	9	10
set_1	0	1	4	8	9
Other	0	0	1	3	11
$Order_2$	i=10	i=20	i=30	i=40	i=50
set_4	8	10	10	10	10
set_3	2	7	9	10	10
set_2	0	3	6	9	10
set_1	0	0	4	7	10
Other	0	0	1	4	10
$Order_3$	i=10	i=20	i=30	i=40	i=50
set_4	8	9	10	10	10
set_3	2	7	9	10	10
set_2	0	3	7	8	10
set_1	0	0	3	7	9
Other	0	1	1	5	11

Three baseline methods are performed on the same data. Fig. 9 shows the average match ratios of Table 2 and the match ratios of the three baseline methods. The angle-based interest model has weak advantages under different values of i. Table 2 and Fig. 9 demonstrate that the angle-based interest model can effectively recommend texts according to the short-term object interest. On the other hand, in Table 2, most values in one column satisfy a decreasing trend which means the node match degree with short-term weight decreases when time lasts. It matches the physical meaning of the node match degree.

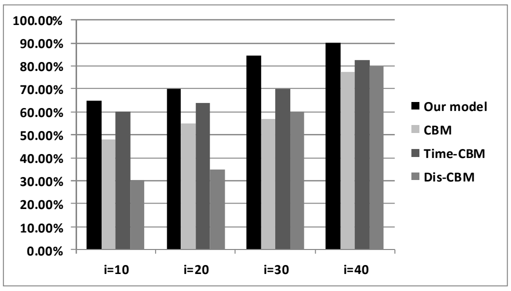


Fig. 9. The average match ratios of the three orders in Table 2 and the match ratios of the three baseline methods.

D.1.2 Experiment on long-term object interest

This experiment is to demonstrate that the angle-based interest model can recommend texts according to the node match degree with long-term weight.

From *Data-Test* 1, 32 topics are chosen and are divided into 4 sets — set_1, set_2, set_3, set_4. Each set has 8 topics. 9 texts are chosen from one topic. So 72 texts are chosen from one set (total 288 texts from the 32 topics). Each topic in the 32 topics has one text rest. $Other_{32}$ denotes the rest 32 texts. Beside the 32 chosen topics, there are 18 rest topics in *Data-Test* 1. 5 target topics are chosen from the 18 topics. In each target topic, the 10 texts are divided into 5 groups. One group has 2 texts. $g_{i,j}$ denotes the j^{th} group in i^{th} target topic (1<=i<=5, 1<=j<=5). The sets and the groups form four partial reading histories as following (*PRH* stands for partial reading history):

PRH_1= *RandomOrder*(set_1, $g_{1,1}$, $g_{2,1}$, $g_{3,1}$, $g_{4,1}$, $g_{5,1}$),
PRH_2= *RandomOrder*(set_2, $g_{1,2}$, $g_{2,2}$, $g_{3,2}$, $g_{4,2}$, $g_{5,2}$),
PRH_3= *RandomOrder*(set_3, $g_{1,3}$, $g_{2,3}$, $g_{3,3}$, $g_{4,3}$, $g_{5,3}$),
PRH_4= *RandomOrder*(set_4, $g_{1,4}$, $g_{2,4}$, $g_{3,4}$, $g_{4,4}$, $g_{5,4}$).

Function *RandomOrder*() arranges the input texts into a random sequence. Then we merge the four partial reading histories into one whole reading history one after another as: $PRH_1 \rightarrow PRH_2 \rightarrow PRH_3 \rightarrow PRH_4$. The main objects in the 5 target topics are distributed in the whole reading history. So the objects in the 5 target topics are the long-term interested objects. An interest graph is obtained by scanning the whole reading history. According to the long-term object interest, the 10 texts from $g_{1,5}$, $g_{2,5}$, $g_{3,5}$, $g_{4,5}$ and $g_{5,5}$, are the ideal results, denoted as *G*. $Other_{32}$ is a comparative group. The texts in $Other_{32}$ are relevant to the whole reading history, but do not meet the long-term object interest because the main objects in $Other_{32}$ concentrate in one partial reading history. $Other_{130}$ denotes 130 texts from the rest 13 topics. The texts in $Other_{130}$ are not relevant to the whole reading history. The node match degree with long-term weight between the interest graph and each text in $G \cup Other_{32} \cup Other_{130}$ is calculated. The experiment is performed three times through rerunning *RandomOrder*() to get different orders of the four partial reading histories. Table 3 shows the experimental results. In Table 3, no matter i = 10 and i = 20, the values in row "*G*" are greater than the values in rows "$Other_{32}$" and "$Other_{130}$". Table 4 shows the average match ratios of Table 3 and the match ratios of the three baseline methods. When i = 10, the ideal set is *G*. The angle-based interest model is a little better than *Dis-CBM* and much better than *CBM* and *time-CBM*. Table 3 and Table 4 demonstrate that the angle-based interest model can effectively recommend texts according to the long-term object interest.

Table 3. The experimental results on the node match degree with long-term weight.

$Order_1$	i=10	i=20
G	7	9
$Other_{32}$	3	9
$Other_{130}$	0	2

(a)

$Order_2$	i=10	i=20
G	8	10
$Other_{32}$	2	8
$Other_{130}$	0	2

(b)

$Order_3$	i=10	i=20
G	8	10
$Other_{32}$	2	9
$Other_{130}$	0	1

(c)

Table 4. The average match ratios of Table 3 and the match ratios from the three baseline methods.

	Our model	*CBM*	*Time-CBM*	*Dis-CBM*
i=10	76.7%	30%	20%	60%

D.2 Experiments on the interest on link structure

Data-Test 2 is used in the experiments. *Data-Test 2* consists of two parts. The texts in $part_1$ are relevant to *George W. Bush* and the texts in $part_2$ are relevant to *Saddam*. $part_1$ is divided into 5 sets respectively on 5 aspects of *Bush* — the education and family background ($set_{1,1}$), campaigning president ($set_{1,2}$), tax reduce policy ($set_{1,3}$), 911 attack and Afghanistan war ($set_{1,4}$), and Iraq war ($set_{1,5}$). $part_2$ is divide into 5 sets respectively on 5 aspects of *Saddam* — the young ($set_{2,1}$), rise to power ($set_{2,2}$), political and cultural image ($set_{2,3}$), Iran-Iraq war ($set_{2,4}$), and Iraq war ($set_{2,5}$). $set_{i,j}$ denotes the *j*th set of part *i*. Each set in $part_1$ and $part_2$ contains 15 texts. $text_{i,j,k}$ denotes the *k*th text in the *j*th set of part *i*. *i* is an integer within [1, 2]; *j* is an integer within [1, 5]; *k* is an integer within [1, 15]. $set_{1,5}$ and $set_{2,5}$ describe the same event and thusly match similar link structures. Then we merge $part_1$ and $part_2$ into one part by putting together $text_{1,j,k}$ and $text_{2,j,k}$ into one text. $text_{2,j,k}$ is after $text_{1,j,k}$. (*j* changes from 1 to 5 and *k* changes from 1 to 15). 5 new sets of texts are obtained after merging — set_1, set_2, set_3, set_4 and set_5. Each set contains 15 texts and each text is relevant to *Bush* and *Saddam*. In the following experiments, set^{m}_{n} denotes *m* texts randomly chosen from set_n.

D.2.1 Experiment on short-term link interest

This experiment is to demonstrate that the angle-based interest model can recommend texts according to the edge match degree with short-term weight.

A reading history with 80 texts is built from *Data-Test* 2 as following: $Other_{50} \rightarrow RandomOrder(set^{10}_{3}) \rightarrow RandomOrder(set^{10}_{4}) \rightarrow RandomOrder(set^{10}_{5})$. The texts from set^{10}_{5} are most close to the current time. $Other_{50}$ denotes 50 texts from the Internet that are irrelevant with *Data-Test* 2 and the 50 texts are irrelevant with each other. An interest graph is obtained by scanning the reading history.

A testing set (total 45 texts) is built that contains (1) set^{5}_{3}, set^{5}_{4} and set^{5}_{5} which respectively denote the rest texts from set_3, set_4 and set_5 (total 15 texts), (2) the 15 texts in set_1 and (3) the 15 texts in set_2. *i* texts are recommended from the testing set according to the short-term link interest. When $i = 5$, the ideal set is set^{5}_{5}; when $i = 10$, the ideal set is $set^{5}_{5} \cup set^{5}_{4}$; when $i = 15$, the ideal set is $set^{5}_{5} \cup set^{5}_{4} \cup set^{5}_{3}$.

To demonstrate that the node match degree with short-term weight is not suitable for recommending texts according to the short-term link interest, the node match degree with short-term weight between each text in the testing set and the interest graph is calculated. The texts with the first *i* node match degrees are recommended (i = 5, 10, 25). The results are shown in Table 5 (a) which are quite different from the ideal results. The reason is that the short-term weights of the two nodes — "*Bush*" and "*Saddam*" — are much higher than other nodes because they appear many times recently. Any text containing "*Bush*" and "*Saddam*" many times has high node match degree with short-term weight. So the texts in set_1 or set_2 have high node match degrees with short-term weight, but the texts do not meet the short-term link interest.

Table 5. The experimental results on the edge match degree with short-term weight.

	i=5	*i*=10	*i*=15
set_1	0	1	2
set_2	1	2	2
set^{5}_{3}	1	1	3
set^{5}_{4}	2	3	4
set^{5}_{5}	1	3	4

(a)

$Order_1$	*i*=5	*i*=10	*i*=15
set_1	0	0	1
set_2	0	0	0
set^{5}_{3}	0	2	5
set^{5}_{4}	1	3	4
set^{5}_{5}	4	5	5

(b)

$Order_2$	*i*=5	*i*=10	*i*=15
set_1	0	0	0
set_2	0	0	1
set^{5}_{3}	0	2	4
set^{5}_{4}	2	4	5
set^{5}_{5}	3	4	5

(c)

$Order_3$	*i*=5	*i*=10	*i*=15
set_1	0	0	0
set_2	0	1	1
set^{5}_{3}	0	2	5
set^{5}_{4}	1	2	4
set^{5}_{5}	4	5	5

(d)

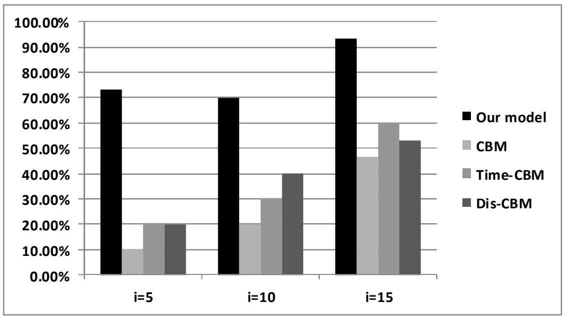


Fig. 10. The average match ratios of the three orders in Table 5 and the match ratios of the three baseline methods.

The edge match degrees with short-term weight between the texts in the testing set and the interest graph are calculated. The texts with the first *i* (*i* = 5, 10, 25) edge match degrees are recommended. The experiment is performed three times by rerunning *RandomOrder*(). The orders of texts in one set are different and the order of sets (set^{10}_{3}→set^{10}_{4}→set^{10}_{5}) remains stable. The results are shown in Table 5 (b), (c), (d). The results satisfy that the most recommended texts belong to the ideal sets when *i* = 5, 10, 15. So the angle-based interest model can effectively recommend the texts matching the short-term link interest.

Fig. 10 shows the average match ratios of Table 5 (b), (c), (d) and the match ratios of the three baseline methods. The match ratios from the angle-based interest model are much better than the baseline methods. So the angle-based interest model is much more effective than the baseline methods for recommending text according to the short-term link interest.

D.2.2 Experiment on long-term link interest

This experiment is to demonstrate that the angle-based interest model can recommend texts according to the edge match degree with long-term weight.

set_6 denotes 150 texts from the Internet that are irrelevant with *Data-Test* 2 and the 150 texts are irrelevant with each other. Five partial reading histories are obtained as following:

PRH_1= *RandomOrder* (set^{20}_{6}, set^{5}_{1}, set^{2}_{5}),
PRH_2= *RandomOrder* (set^{20}_{6}, set^{10}_{2}, set^{2}_{5}),
PRH_3= *RandomOrder* (set^{20}_{6}, set^{10}_{3}, set^{2}_{5}),
PRH_4= *RandomOrder* (set^{20}_{6}, set^{5}_{1}, set^{2}_{5}),
PRH_5= *RandomOrder* (set^{20}_{6}, set^{10}_{4}, set^{2}_{5}).

set^{2}_{5}, set^{5}_{1} and set^{20}_{6} denote different sets of texts in different partial reading histories. Then the five partial reading histories are merged into one whole reading history as $PRH_1 \rightarrow PRH_2 \rightarrow PRH_3 \rightarrow PRH_4 \rightarrow PRH_5$. The texts in PRH_5 are most close to the current time. An interest graph is obtained by scanning the whole reading history. A testing set is build that contains the rest 50 texts from set_6 and the rest 25 texts from $set_1 \cup set_2 \cup set_3 \cup set_4 \cup set_5$. According to the whole reading history, the texts in set_5 match the long-term interested link structure because the texts in set_5 are most evenly distributed. Compared with set_2, set_3 and set_4, the texts in set_1 match the long-term interested link structure more close because the texts in set_1 appear in PRH_1 and PRH_4. *i* texts are recommended from the testing set. When *i* = 5, the ideal set is the 5 texts from set_5 in the testing set; when *i* = 10, the ideal set is the 10 texts from $set_5 \cup set_1$ in the testing set, when *i* = 25, the ideal set is the 25 texts from $set_5 \cup set_4 \cup set_3 \cup set_2 \cup set_1$ in the testing set.

To demonstrate that the node match degree with long-term weight is not suitable for recommending texts according to the long-term link interest, the node match degree with long-term weight between each text in the testing set and the interest graph is calculated. The texts with the first *i* node match degrees are recommended (*i* = 5, 10, 25). The results are shown in Table 6 (a) which are quite different from the ideal results. The reason is that the long-term weights of the two nodes — “*Bush*” and “*Saddam*” — are much higher than other nodes. Any text containing “*Bush*” and “*Saddam*” many times has high node match degree with long-term weight.

Table 6. The experimental results on the edge match degree with long-term weight.

	i=5	i=10	i=25
set_5	2	3	5
set_4	1	2	4
set_3	0	2	5
set_2	1	1	4
set_1	1	2	5
set_6	0	1	2

(a)

$Order_1$	i=5	i=10	i=25
set_5	4	5	5
set_4	0	0	5
set_3	0	2	5
set_2	0	0	4
set_1	1	3	5
set_6	0	0	1

(b)

$Order_2$	i=5	i=10	i=25
set_5	5	5	5
set_4	0	1	5
set_3	0	0	4
set_2	0	1	4
set_1	0	3	5
set_6	0	0	2

(c)

$Order_3$	i=5	i=10	i=25
set_5	4	5	5
set_4	0	0	5
set_3	0	1	5
set_2	0	0	5
set_1	1	4	5
set_6	0	0	0

(d)

The edge match degrees with long-term weight between the texts in the testing set and the interest graph are calculated. The texts with the first i (i = 5, 10, 25) edge match degrees are recommended. The experiment is performed three times by rerunning *RandomOrder*(). The order of partial reading histories remains stable. The results are shown in Table 6 (b), (c), (d). When i = 5, the most recommended texts belong to set_5. When i = 10, the most recommended texts belong to $set_5 \cup set_1$. When i = 25, the most recommended texts belong to $set_1 \cup set_2 \cup set_3 \cup set_4 \cup set_5$. The results demonstrate that the angle-based interest model can effectively recommend texts according to the long-term link interest.

Fig. 11 shows the average match ratios of Table 6 (b), (c), (d) and the match ratios of the three baseline methods. When i = 25, the four methods work well because the recommended texts from the four methods are relevant to the reading history. When i = 5 or 10, the match ratios from the angle-based interest model are much better than the baseline methods. So the angle-based interest model is more effective than the baseline methods for recommending text according to the long-term link interest.

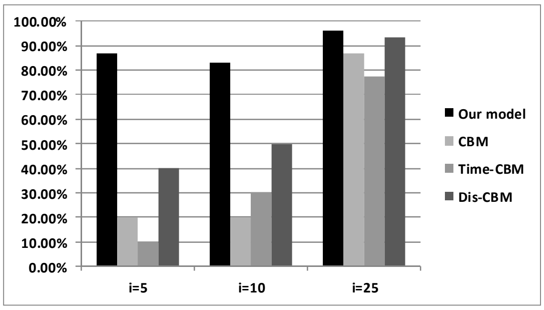


Fig. 11. The average match ratios of the three orders in Table 6 and the match ratios of the three baseline methods.

Appendix E. Experiments on complex angles

E.1 Experiment on the angle that concerns the new information on the short-term interested objects

Data-Test 1 is used. 12 topics are randomly chosen from the 50 topics. T^n_m denotes the n texts chosen from the m^{th} topic. T_m denotes the texts from m^{th} topic. Four partial reading histories are obtained as following,

PRH_1 = *RandomOrder*(T^5_1, T^5_2),

PRH_2 = *RandomOrder*(T^5_3, T^5_4, T^1_9, T^1_{10}, T^2_{11}),

PRH_3 = *RandomOrder*(T^5_5, T^5_6, T^1_9, T^1_{10}, T^2_{12}),

PRH_4 = *RandomOrder*(T^5_7, T^5_8, T^5_9, T^5_{10}, T^5_{11}, T^5_{12}).

Once a text appears in one partial reading history, it does not appear in another partial reading history. In the partial reading histories, T_9 and T_{10} are distributed in PRH_2, PRH_3 and PRH_4; T_{11} is distributed in PRH_2

and PRH_4; T_{12} is distributed in PRH_3 and PRH_4. The partial reading histories are merged into one whole reading history as: $PRH_1 \rightarrow PRH_2 \rightarrow PRH_3 \rightarrow PRH_4$. The texts in PRH_4 are most close to the current time. An interest graph is obtained by scanning the whole reading history. The main objects in T_7 and T_8 are recently interested objects and do not appear before. So the texts in T_7 and T_8 meet the angle that concerns the new information on the short-term interested objects.

From T_1 to T_8, 3 texts are randomly chosen from the rest texts of each topic (total 24 texts). And each topic from T_9 to T_{12} has 3 rest texts (12 texts). The 36 (24+12) texts form a testing set. According to the combination of the match degrees of the angle that concerns the new information on the short-term interested objects, i (i = 6, 12) texts are recommended from the testing set. Three different coefficient vectors are set as shown in Table 7 (a), (b) and (c). The experiment is preformed three times with three different orders of the partial reading histories. The average values of the results on the three orders are calculated. There are three expectations.

(1). If the coefficient vector is the default value in Table 7 (a), the results should mostly belong to T_7 and T_8 because the texts in T_7 and T_8 meet the angle most closely. Except T_7 and T_8, the results should mostly belong to T_5, T_6 and T_{12} because the texts in the three topics meet the angle second most closely.

(2). If a user sets the coefficient vector as CV_1 in Table 7 (b), the results should be more relevant to the objects in PRH_4. So, the results, except T_7 and T_8, should be more distributed in T_9-T_{12} compared with Table 7 (a). CV_1 means the user prefers "*short-term interested*" to "*new*".

(3). If a user sets the coefficient vector as CV_2 in Table 7 (c), the results should be more relevant to the objects which are not long-term interested in. So, the results, except T_7 and T_8, should be more distributed in T_1-T_6 compared with Table 7 (a). CV_2 means the user prefers "*new*" to "*short-term interested*".

The experimental results in Table 7, especially the average values, match the three expectations. So the angle-based interest model is effective to handle the angle that concerns the new information on the short-term interested objects. And the recommended texts correctly change with the coefficient vector.

Table 7. The experimental results on the angle that concerns the new information on the short-term interested objects

(a) CV=(0.5, 0.5)	$Order_1$		$Order_2$		$Order_3$		AVG	
i=	6	12	6	12	6	12	6	12
T_1 —T_4	0	0	0	0	0	0	0	0
T_5, T_6	1	3	0	3	1	2	0.67	2.67
T_7, T_8	4	6	5	6	4	6	4.33	6
T_9, T_{10}	0	0	0	0	0	0	0	0
T_{11}	0	1	0	1	0	1	0	1
T_{12}	1	2	1	2	1	3	1	2.33
(b) CV_1=(0.3, 0.7)								
T_1 —T_4	0	0	0	0	0	0	0	0
T_5, T_6	0	1	0	0	0	0	0	0.33
T_7, T_8	3	6	4	6	3	6	3.33	6
T_9, T_{10}	0	1	0	2	0	3	0	2
T_{11}	1	2	0	2	1	1	0.67	1.67
T_{12}	2	2	2	2	2	3	2	2.33
(c) CV_2=(0.7, 0.3)								
T_1 —T_4	0	2	0	2	0	1	0	1.67
T_5, T_6	1	3	0	4	1	4	0.67	3.67
T_7, T_8	5	6	6	6	5	6	5.33	6
T_9, T_{10}	0	0	0	0	0	0	0	0
T_{11}	0	0	0	0	0	0	0	0
T_{12}	0	1	0	0	0	1	0	0.67

E.2 Experiment on the angle that concerns the new link structure containing the long-term interested objects

Data-Test 2 is used. A whole reading history is formed by merging following partial reading histories one after another. set_6 denotes 100 texts from the Internet which are irrelevant with set_k (k = 1, 2, 3, 4, 5). And the texts in set_6 are irrelevant with each other.

PRH_1= RandomOrder (set^{10}_6, set^{3}_1), PRH_2= RandomOrder (set^{10}_6, set^{5}_2), PRH_3= RandomOrder (set^{10}_6, set^{3}_1), PRH_4= RandomOrder (set^{10}_6, set^{10}_3), PRH_5= RandomOrder (set^{10}_6, set^{5}_2), PRH_6= RandomOrder (set^{10}_6, set^{4}_1), PRH_7= RandomOrder (set^{10}_6, set^{10}_4).

set^{10}_6, set^3_1, set^5_2 denotes different sets of texts in different *PRH*s. Because the texts of set_5 do not appear in the reading history and the main objects in set_1, set_2, set_3, set_4, set_5 and set_6 are similar, the link structure in set_5 is neither short-term nor long-term and contains the long-term interested objects. So the texts in set_5 meet the angle that concerns the new link structure containing the long-term interested objects.

Except the texts in the whole reading history, each set from set_1 to set_4 has 5 rest texts (total 20 texts), plus 10 texts from set_5 and the 30 rest texts from set_6, the 60 texts form a testing set. The first i (i = 10, 20) texts are recommended from the testing set.

In Table 8 (a), the coefficient vector is set as the default value. The ideal results are: (1) If i = 10, the ideal results are the 10 texts in set_5; (2) If i = 20, the recommended texts belong to set_5 and set_3. The reason is that the texts in set_3 satisfy "*Low edge match degree with short-term weight*" and "*Low edge match degree with long-term weight*" more close than the texts in other sets. The experiment is performed three times with three different orders of the partial reading histories. The results are shown in Table 8 and match the ideal results.

Then we give different coefficient vectors as CV_1 = (0.2, 0.6, 0.2) and CV_2 = (0.2, 0.2, 0.6). In Table 8 (b), CV_1 emphasizes "*Low edge match degree with short-term weight*" which means the user wants to ensure the recommended texts do not meet the short-term link interest; CV_2 emphasizes "*Low edge match degree with long-term weight*" which means the user wants to ensure the recommended texts do not meet the long-term link interest. So the recommended texts should be more distributed in set_3 and set_5 while setting CV_1, and the recommended texts should be more distributed in set_4, set_3 and set_5 while setting CV_2. Table 8 (b) and (c) show the experimental results which match the above expectations. Therefore the angle-based interest model is effective to handle the angle that concerns the new link structure containing the long-term interested objects. And the recommended texts correctly change with the coefficient vector.

Table 8. The experimental results on the angle that concerns the new link structure containing the long-term interested objects

(a) CV = (0.33, 0.33, 0.33)	$Order_1$		$Order_2$		$Order_3$		AVG	
i=	10	20	10	20	10	20	10	20
set_1	0	0	0	0	0	0	0	0
set_2	1	3	0	1	0	1	0.33	1.67
set_3	1	5	2	4	1	6	1.33	5
set_4	0	2	0	4	0	3	0	3
set_5	8	10	8	10	9	10	8.33	10
set_6	0	0	0	1	0	0	0	0.33
(b) CV_1 = (0.2, 0.6, 0.2)								
set_1	0	0	0	0	0	0	0	0
set_2	0	3	0	3	1	1	0.33	2.33
set_3	4	7	3	6	4	8	3.67	7
set_4	0	0	0	0	0	0	0	0
set_5	6	10	7	10	5	10	6	10
set_6	0	0	0	1	0	1	0	0.67
(c) CV_2 = (0.2, 0.2, 0.6)								
set_1	0	0	0	0	0	0	0	0
set_2	0	0	0	0	0	0	0	0
set_3	2	4	2	4	1	5	1.67	4.33
set_4	2	6	3	6	2	4	2.33	5.33
set_5	6	10	5	9	7	10	6	9.67
set_6	0	0	0	1	0	1	0	0.67

Appendix F. Experiments on coefficients

F.1 Experiment on the descending function in calculating the short-term weight

This experiment is to evaluate the performance of the angle-based interest model while setting different λ in equation (8) and compares $\Omega(k)$ with other typical functions. λ is tested from 1.0 to 4 step 0.1. The values of λ out of the range lead to low and unstable performance. λ is set within different areas as shown in Fig. 12. For each value of λ, one match ratio on the node match degree with short-term weight and one match ratio on the edge match degree with short-term weight are obtained through conducting the two experiments on short-term object interest and short-term link interest. The results are shown in Fig. 12. When 1.1 < λ <= 1.2, the results are better than others. But there is not a certain value of λ in [1.1, 1.2] that can always achieve best performance while facing different orders of the partial reading histories.

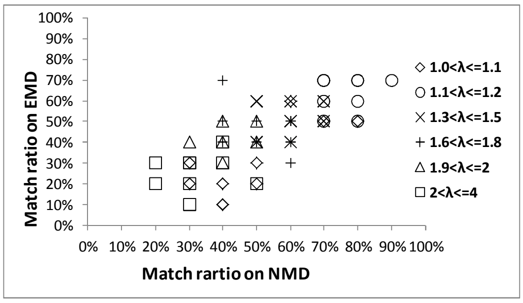


Fig. 12. The match ratios by setting different λ. The match ratios on the node match degree with short-term weight come from the experiment on short-term object interest. The match ratios on the edge match degree with short-term weight come from the experiment on short-term link interest.

Two typical functions are compared with $\Omega(k)$. One is a linear function: 1-(k-1)/$SIZE$; the other is an exponential function: $e^{-(k-1)}$. Fig. 13 shows the experimental results that demonstrate $\Omega(k)$ is better than the two functions.

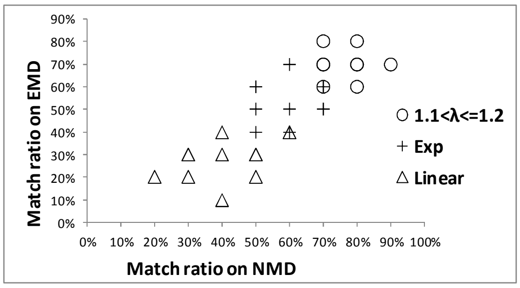


Fig. 13. The match ratios on the node match degree with short-term weight and the edge match degree with short-term weight by using different functions. The horizontal axis represents the match ratios on the node match degree with short-term weight. The vertical axis represents the match ratios on the edge match degree with short-term weight.

F.2 Experiment on the coefficients in calculating the distributing coefficient

The experiment is to test the performance of the angle-based interest model while setting different m, n and α in equation (10). Because α does not influence the comparison between the arrays on even distribution, m and n are firstly tested.

The *ITA*/*IFA* array in the calculation of the distributing coefficient can be equivalently considered as 0-1 array. 300 0-1 arrays are randomly generated as a group of testing data. Each array contains n_c components. If two arrays can hardly be intuitively compared because of the similar distribution, one of the two arrays is removed. *number_remain* arrays are obtained and get an ideal rank by intuitively comparing the even distribution of the arrays. On the other hand, a rank is obtained by calculating and comparing the distributing coefficients of the *number_remain* arrays. Given two arrays c and d, if c is before d in the ideal rank and c is after d in the calculated rank, or the opposite, then (c, d) is a wrong pair. Assuming the number of wrong pairs is w_p, the match ratio of the distributing coefficient (MR_{DC}) is defined as equation (15),

$$MR_{DC} = \frac{w_p \times 2}{(number\_remain - 1) \times number - remain} \qquad (15).$$

For certain m and n, the process is performed 6 times (First time, n_c = 50, second time, n_c = 100, third time n_c = 150, 4th time n_c = 200, 5th time n_c = 250, 6th time n_c = 300) to get the average MR_{DC} as shown in Fig. 14. m

and *n* change from 1 to 8 step 1. When *m*=3 and *n*=2, MR_{DC} is better than others. When *m* <= *n*, MR_{DC} are relatively low.

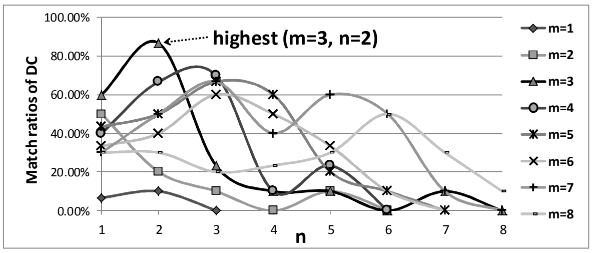


Fig. 14. The average match ratios of the distributing coefficient when *m* and *n* change. The vertical axis represents the match ratios of the distributing coefficient. The horizontal axis represents *n*. Each curve represents a value of *m*.

Given *m*=3 and *n*=2, different values of *α* are tested as shown in Fig. 15. When *α* is around 0.6, the match ratios of the node match degree with long-term weight and the match ratios of the edge match degree with long-term weight are relatively high. The horizontal axis represents *α*. The vertical axis represents the match ratios while performing the two experiments on long-term object interest and long-term link interest.

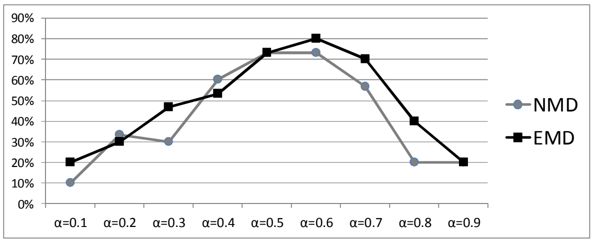


Fig. 15. The match ratios when *α* changes. The match ratios on the node match degree with long-term weight come from the experiment on long-term object interest. The match ratios on the edge match degree with long-term weight come from the experiment on long-term link interest.