Future Generation Computer Systems 64 (2016) 211–226

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

An angle-based interest model for text recommendation

Bei Xu *, 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

HIGHLIGHTS

- The model represents the multiple basic angles of reading interest.
- The basic angles reflect the pattern and persistence of reading interest.
- The model represents a complex angle by combining basic angles.
- Different angles correspond to different sets of required texts.

ARTICLE INFO

Article history:
Received 18 March 2015
Received in revised form
13 April 2016
Accepted 16 April 2016
Available online 9 May 2016

Keywords:
Text recommendation
Interest model
Multi-angle interest

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 interest. 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 in object and the interest in the link structure of objects. Experiments with news-scale text data show that the interest in object and the interest in link structure have real requirements, and it is effective to recommend texts according to the angles.

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 [1–4]. Keywords are extracted from text or provided by users, e.g., in iGoogle. Keywords are often calculated with weights [5], e.g. *tf-idf* [6–8]. Relevant texts are recommended through the techniques on words' weights, e.g. using cosine [6] or collaborative-filtering [9–11]. Keywords can also be organized into topics to select the texts close to the interested topics [12,13]. There are other kinds of keyword methods. Histogram methods use histogram to analyze the statistics of keywords [7,14]. 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 [15,16], and then use collaborative filtering to analyze user evaluations [17].

The background knowledge methods incorporate concept hierarchies, ontology or encyclopedic knowledge to analyze a user's interests and help users express requirements [18–21]. 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 [22].

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) [4,23]. 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 the 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

* Corresponding author at: Institute of Cyber–Physical–Social Intelligence, Nanjing University of Posts and Telecommunications, China.
E-mail address: xubei@njupt.edu.cn (B. Xu).

http://dx.doi.org/10.1016/j.future.2016.04.011
0167-739X/

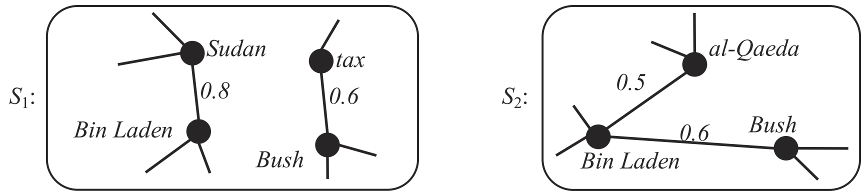

Fig. 1. Two partial link structures interested by two users, where a node denotes an object and an edge denotes that the two objects are linked to a certain degree.

domains that they are not interested in [24]. 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 [25,26,8]. 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 [27]. 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 9/11-attack. Neither of the two parts in h_1 directly relate to 9/11-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.

The following is a short text related to 9/11-attack from *CNN*. It meets u_1's interest and u_2's interest in object because it shares a lot of objects with h_1 and h_2. But it does not meet u_1's interest in 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 in link structure because its structure and S_2 have intersection of link.

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 [7,28,8]. The components in a vector are words or phrases and the weights of the components usually are *tf-idf* values. The topic model analyzes the term distributions (e.g., *PLSA* 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 [29].

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 [30,25]. Some methods predicate user behaviors probabilistically based on the user's historical behaviors (e.g., click distribution) [31,4,32]. 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 [33,34].

The context-aware text recommendation considers the context information [35], e.g., time or location [36]. 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 in 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 [27,37]. 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 classify 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 [38]. 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 interest in some objects usually searches 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).