Interest Representation, Enrichment, Dynamics, and Propagation: A Study of the Synergetic Effect of Different User Modeling Dimensions for Personalized Recommendations on Twitter

Guangyuan Piao(✉) and John G. Breslin

Insight Centre for Data Analytics, NUI Galway, IDA Business Park, Lower Dangan, Galway, Ireland
guangyuan.piao@insight-centre.org, john.breslin@nuigalway.ie

Abstract. Microblogging services such as Twitter have been widely adopted due to the highly social nature of interactions they have facilitated. With the rich information generated by users on these services, user modeling aims to acquire knowledge about a user's interests, which is a fundamental step towards personalization as well as recommendations. To this end, researchers have explored different dimensions such as (1) *Interest Representation*, (2) *Content Enrichment*, (3) *Temporal Dynamics* of user interests, and (4) *Interest Propagation* using semantic information from a knowledge base such as DBpedia. However, those dimensions of user modeling have largely been studied separately, and there is a lack of research on the synergetic effect of those dimensions for user modeling. In this paper, we address this research gap by investigating 16 different user modeling strategies produced by various combinations of those dimensions. Different user modeling strategies are evaluated in the context of a personalized link recommender system on Twitter. Results show that *Interest Representation* and *Content Enrichment* play crucial roles in user modeling, followed by *Temporal Dynamics*. The user modeling strategy considering *Interest Representation*, *Content Enrichment* and *Temporal Dynamics* provides the best performance among the 16 strategies. On the other hand, *Interest Propagation* has little effect on user modeling in the case of leveraging a rich *Interest Representation* or considering *Content Enrichment*.

1 Introduction

With the popularity of microblogging services such as Twitter[1], the amount of information available on the Social Web is increasing exponentially. While this information is a valuable resource, its sheer volume limits its value [9]. On the Social Web, as the amount of information available causes information overload for users, the demand for personalized approaches towards information consumption increases. User (interest) modeling aims to analyze user activities on the

[1] https://www.twitter.com.

E. Blomqvist et al. (Eds.): EKAW 2016, LNAI 10024, pp. 496–510, 2016.
DOI: 10.1007/978-3-319-49004-5_32

Table 1. A sample tweet posted by Bob [22]

My Top 3 #lastfm Artists: Eagles of Death Metal(14), The Black Keys(6) & The Wombats(6). http://www.last.fm/user/bob

Social Web in order to provide personalized services for users. To create qualitative and quantitative user models for microblogging services such as Twitter, several design dimensions have been investigated in previous studies.

Interest Representation. The first step of user modeling is to determine how to represent user interests. Several approaches such as *bag-of-words*, *topic models* or *bag-of-concepts* have been used for representing user interests. Take an example from our own recent work (see Table 1 [22]), by using the bag-of-concepts approach, we can assume that the user is interested in DBpedia[2] entities such as `dbpedia`[3] and `dbpedia:The_Wombats` based on a tweet posted by a user named Bob. In addition, we can exploit background knowledge of entities from a Knowledge Base (KB) for extending user interests, e.g., categories of the entities in DBpedia. Throughout the paper, by a *concept* we mean an *entity*, *category* or *class* from a KB (e.g., DBpedia) for representing user interests.

Content Enrichment. As the ideal length of User-Generated Content (UGC) on microblogging services is short[4], there is a need to enrich this short content to better understand the context of it. Embedded links (URLs) in a tweet can be used to enrich the short content, and provide additional information about the tweet. For example, we can follow the link in the sample tweet to retrieve more information about Bob's musical interests. Many sources have shown that a large portion of tweets and retweets contain links[5,6].

Temporal Dynamics. Users might be interested in different topics over time. To capture the dynamics of user interests, some previous studies have used short-term profiles (e.g., considering a user's activities during the last two weeks only), while others have proposed interest decay functions to discount older interests.

Interest Propagation. This dimension exploits cross-domain background knowledge about concepts from a KB such as DBpedia. Based on the concepts directly spotted from UGC, related concepts in the KB can be leveraged for enriching user interest profiles. For instance, Bob (see Table 1) might be interested in `dbpedia:Indie_rock` as he likes indie rock artists such as `dbpedia:The_Black_Keys` and `dbpedia:The_Wombats` based on background knowledge from DBpedia, e.g., `dbpedia:The_Black_Keys` → `dbpedia-owl`[7]`:genre` → `dbpedia:Indie_rock`. Throughout the paper, we

[2] http://wiki.dbpedia.org.
[3] The prefix `dbpedia` denotes http://dbpedia.org/resource/:The_Black_Keys.
[4] http://goo.gl/uewQLu.
[5] http://marketingrelevance.com/news/04/tweet-interesting-information/.
[6] http://goo.gl/RGC16n.
[7] The prefix `dbpedia-owl` denotes http://dbpedia.org/ontology/.

denote the concepts that can be directly extracted from a user's tweets as ***primitive interests*** (e.g., `dbpedia:The_Wombats`), and the concepts that can be propagated from those primitive interests as ***propagated interests*** (e.g., `dbpedia:Indie_rock`).

Although related work reveals many promising insights with respect to those user modeling dimensions, there exists little research on studying the synergetic effect achieved by considering those dimensions together [20]. As those dimensions are not necessarily exclusive of each other, this has in turn motivated us to implement a user modeling framework which can exploit different dimensions at the same time for generating user interest profiles. We then evaluate different user interest profiles generated by different user modeling strategies in the context of a personalized link (URL) recommender system on Twitter.

The contributions of this work are summarized as follows.

- We implemented a user modeling framework, which can incorporate different combinations of four dimensions: (1) *Interest Representation*, (2) *Content Enrichment*, (3) *Temporal Dynamics*, and (4) *Interest Propagation*, to investigate (how) can we combine these different dimensions to retrieve better user interest profiles. To our knowledge, this is the first comprehensive study on these four dimensions.
- We evaluate 16 user modeling strategies generated by different combinations of methods for those four dimensions in the context of link recommendations on Twitter using four different evaluation metrics.

The organization of the rest of the paper is as follows. Section 2 gives some related work, and Sect. 3 describes our user modeling framework. In Sect. 4, we present the experiment setup for our study. Experiment results are presented in Sect. 5. Finally, Sect. 6 concludes the paper with some future work.

2 Related Work

In this section, we provide an overview of some related work from the literature for the aforementioned dimensions in user modeling.

Representation of User Interests. To represent user interest profiles, researchers began by *word-based* approaches such as *bag-of-words* [8,17], *topic modeling* [10]. Degemmis et al. [8] proposed a specific *word-based* approach - using WordNet[8] synsets (which are unordered sets of synonyms) for representing user interests. They showed that their *bag-of-synsets* approach outperformed a *bag-of-words* approach. As *word-based* approaches focus on the words themselves and do not provide semantic information about the words or the relationships among them, a research direction has been proposed over the past few years that uses *concept-based* representations of user interests using a KB from Linked Data form (e.g., Freebase, DBpedia) [4,5,19,23] or using an encyclopedia such as Wikipedia [12,15,16,18]. More recently, we showed that using synsets and

[8] https://wordnet.princeton.edu/.

concepts together for representing user interests can improve the quality of user modeling on Twitter in the context of link recommendations [21].

Enrichment for Short Messages. To better understand the semantics of short messages generated in microblogging services such as Twitter, some researchers have used the content of embedded links (URLs) in short messages to enrich the content [4,13]. In [4], the authors first used URLs in a user's tweets to enrich their content. After that, the user's interest profiles were constructed based on the enriched content. They showed that enriching short content for retrieving user interests enhances the variety and quality of the generated user profiles, and improves the performance of news recommendations.

Dynamics of User Interests. Many methods have been proposed to incorporate the temporal dynamics of user interests based on the hypothesis that the interests of users change over time [2,3,7,19]. For example, Abel et al. [3] studied short-term and long-term user profiles from Twitter for news recommendations. To construct a short-term user profile for a given user, they only used the user's tweets within the last two weeks. On the other hand, a long-term user profile was generated based on the user's entire historical tweets. Another line of work [2,7,19] that incorporates temporal dynamics applies a decay function to the interests of users. The rationale behind the decay function is that higher weights should be given to interests that have occurred recently and lower weights given to older interests.

Interest Propagation using Background Knowledge. There are various related works [19,22,23] that enrich *concept-based* user interest profiles using background knowledge. In [19], the authors built *category-based* user interest profiles by exploring DBpedia categories of entities, e.g., using categories such as `dbc`[9]`:Apple_Inc._executives` to denote user interests if a user is interested in `dbpedia:Steve_Jobs`. Piao et al. [22] proposed a mixed approach that combines the entity- and category-based profiles with the discounting strategy from [19], and proved that the mixed approach performs better than either the entity- or category-based approach. Building on this in a later work [23], the authors showed that by using Concept Frequency - Inverse Document Frequency (CF-IDF) as the weighting scheme and by leveraging different types of information from DBpedia to extend user profiles (i.e., *categories*, and *connected entities via different properties*), the quality of user modeling can be improved.

There are also some studies for user modeling with respect to a specific domain of user interests. For example, Abel et al. [5] proposed using DBpedia to extend user profiles with respect to point of interests (POI), and Nishioka et al. [18] explored different factors of user modeling for modeling user interests with respect to scientific publications in the economic domain. Different from focusing on user interests in a specific domain, our work focuses on user interests extracted from Twitter which are not limited to a specific domain.

While related work reveals several insights regarding each dimension of user modeling, hybrid approaches combining those different dimensions are

[9] The prefix `dbf` denotes http://dbpedia.org/resource/Category:.

considered only to a limited degree. For example, after enriching tweets with the content of embedded links, it would be interesting to explore if interest propagation using background knowledge further improves the quality of user modeling, or if it has little effect or no effect since enough information may already be available from a user's primitive interests.

3 Content-Based User Modeling

In this section, we first introduce user interest profiles as defined in our work, and then present a general process for generating user interest profiles (Sect. 3.1). Subsequently, we provide details of the methods for each of the user modeling dimensions used in the process (Sect. 3.2).

In this work, we use the same definition from [20] to represent the interests of users, which is specified as follows.

Definition 1. *The interest profile of a user $u \in U$ is a set of weighted DBpedia concepts or WordNet synsets, where with respect to a given user u who has an interest $i \in I$, its weight $w(u, i)$ is computed by a certain function w.*

Here, U denotes the set of users, and I denotes the set of concepts in DBpedia and synsets in WordNet, respectively. The weighting scheme $w(u, i)$ measures the importance of a concept with respect to a user. Previous studies showed that using CF-IDF as the weighting scheme provides better performance than using a Concept Frequency (CF) weighting scheme for user modeling in the context of recommender systems [18,23]. Similar to the TF-IDF weighting scheme used in *word-based* user modeling approaches [1], the rationale behind CF-IDF is discounting the weights of concepts appearing frequently in users' interest profiles and increasing the weights of concepts appearing rarely in users' profiles. In the same way, we use the Interest Frequency - Inverse Document Frequency (IF-IDF) as the weighting scheme for our experiments. More formally, it is defined as follows.

- $w_{IF}(u, i) = the\ frequency\ of\ \ i\ in\ a\ user's\ tweets,$
- $w_{IF-IDF}(u, i) = \underbrace{w_{IF}(u, i)}_{IF} \times \underbrace{\log \frac{M}{m_i}}_{IDF}$

where M is the total number of users, and m_i is the number of users interested in a concept/synset i.

3.1 The Process of Generating User Interest Profiles

Figure 1 presents the process of generating user interest profiles for Twitter considering the aforementioned four different user modeling dimensions. The components with dotted lines are options that can be either can be "enabled" or "disabled" for this user modeling. The process has three major steps:

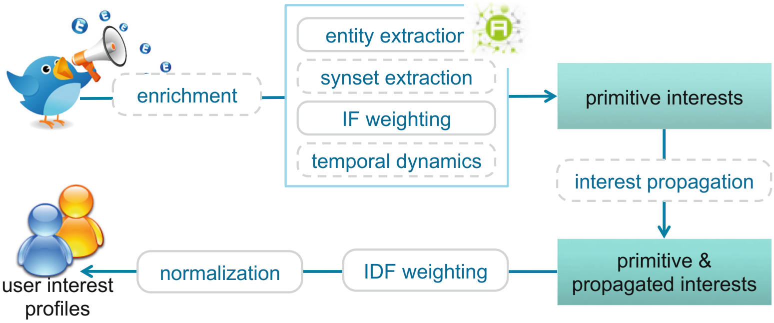


Fig. 1. The process of generating user interest profiles on Twitter

(1) **Primitive interests extraction**. For a given user, we extract all *primitive interests* (DBpedia entities or WordNet synsets) within UGC of a user. If the component *enrichment* is enabled, the content of links embedded in the UGC will also be used for extracting primitive interests.
   - DBpedia entities are extracted using the `Aylien API`[10]. For instance, the API extracts two entities `dbpedia:Microsoft` and `dbpedia:LinkedIn` from the phrase: "*Microsoft to Buy LinkedIn for $26B; LinkedIn to continue as separate brand*". Interest Frequency (IF) is applied to denote the importance of a concept with respect to a user. In addition, it might adhere to strategies for incorporating the *temporal dynamics* of user interests.
   - WordNet synsets can be extracted at the same time as extracting entities. The rationale behinds this is that syntactic information can complement semantic information for generating user interest profiles [21]. For example, given a tweet: "*Just completed a 3.89 km ride. We're gonna need more...*", we can extract synsets such as: s1 = [kilometer, kilometre, km, klick (a metric unit of length equal to 1000 meters (or 0.621371 miles))] and s2 = [drive, ride (a journey in a vehicle (usually an automobile))], which denote the user interests that would be missed if we used a concepts-alone approach.

(2) **Interest propagation.** This component can apply propagation strategies to primitive interests based on background knowledge from DBpedia. The output here is a user interest profile consisting of *primitive interests* as well as *propagated interests*.

(3) **Weighting and normalization**. Finally, the user modeling framework applies Inverse Document Frequency (IDF) to the user interest profile, and further normalizes the profile so that the sum of all weights in the profile is equal to 1: $\sum_{i \in I} w(u,i) = 1$.

[10] http://aylien.com.

Table 2. The design space of user modeling, spanning $2 \times 2 \times 2 \times 2 = 16$ possible user modeling strategies

User modeling dimensions	Interest representation	Content enrichment	Temporal dynamics	Interest propagation
Options	*Concept*	*Enabled*	*Enabled*	*Enabled*
	Synset & concept	*Disabled*	*Disabled*	*Disabled*

Based on the optional components for user modeling (shown with dotted lines in Fig. 1), there are 16 possible strategies which are displayed in Table 2. In the following subsection, we provide details of the methods for each dimension.

3.2 Methods for Each Dimension

Interest Representation: (1) Concept, or (2) Synset & Concept. *Entity recognition* and *synsets extraction* are performed in the first step to extract *primitive interests* from a user's tweets.

Entity recognition in tweets is a challenging task due to the informal nature of and ungrammatical language in tweets. Since our focus in this work is on user modeling and not entity recognition, we have used an existing solution for entity recognition (as does related literature on user modeling).

Different Natural Language Processing (NLP) APIs have been used for DBpedia/Wikipedia entity recognition in the literature. For example, Kapanipathi et al. [12] used the Zemanta API (which is no longer available) after comparing it to other APIs such as DBpedia Spotlight[11], Fattane et al. [24] used tag.me[12], and Piao et al. [23] used the Aylien API, respectively. To better investigate the performance of different APIs, we used the Twitter dataset from [14] which contains annotated 1,603 tweets in total where 1,233 of them contain Wikipedia entities. We tested three different NLP APIs: Aylien API, tag.me and Alchemy API[13], which all provide functionality for extracting entities from a given text and representing these with corresponding DBpedia/Wikipedia URIs. A comparative performance is displayed in Table 3. We opted to use the Aylien API for our experiment since it (1) extracts DBpedia entities (*primitive interests*) identified in tweets, and gives their corresponding

Table 3. Evaluation of NLP APIs for DBpedia/Wikipedia entity recognition

API	Precision	Recall	F-measure
Aylien	0.27	0.26	0.26
Alchemy	0.21	0.17	0.19
tag.me	0.12	0.15	0.14

[11] http://spotlight.dbpedia.org/rest/annotate, the web service was not accessible at the time of writing this paper.

[12] https://tagme.d4science.org/tagme/.

[13] http://www.alchemyapi.com/.

URIs, (2) it has relatively superior performance to the other APIs as shown in Table 3, and (3) it provides 6,900 calls per day, provided on request for research purposes.

Synset extraction is included in the investigation since concepts from a KB could not express user interests completely [21]. On one hand, there might be new concepts/topics emerging in microblogging services such as Twitter, which cannot be found in a KB. On the other hand, the earlier work [21] showed that using WordNet sunsets and DBpedia concepts together is helpful for improving the quality of user interest profiles. In this regard, in the same way from [21], we adopt a method from [8] which extracts WordNet synsets to build *synset-based* user interest profiles.

Content Enrichment: (1) Enabled, or (2) Disabled. We leverage the content of links embedded in a tweet to enrich the original post content. Based on the selected option for the dimension *Interest Representation*, we apply the same extraction method for the content of embedded links. Therefore, in the case of *concepts* being used for *Interest Representation*, the *concepts* extracted from the content of links embedded in tweets will also be considered as user interests if the *Content Enrichment* dimension option is enabled.

Temporal Dynamics: (1) Enabled, or (2) Disabled. In [23], the authors conducted a comparative study on different interest decay functions [2,6,19] in the context of recommender systems on Twitter. Results showed that those functions have similar performance. We choose a variant of the interest decay function from [6], which performed best overall in the comparative study [23]. This decay function [23] measures the expected weight in terms of an interest i for user k at time t by combining three levels of abstractions using a weighted sum as below:

$$w_{ki}^{t} = \mu_{2week} w_{ki}^{t,2week} + \mu_{2month} w_{ki}^{t,2month} + \mu_{all} w_{ki}^{t,all} \tag{1}$$

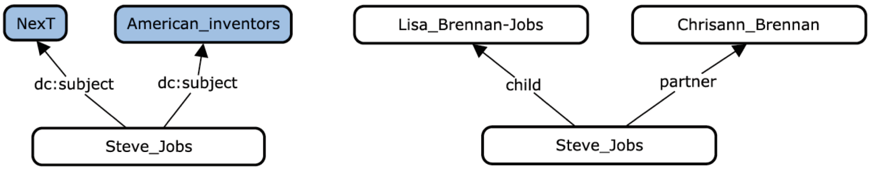
where $\mu_{2week} = \mu$, $\mu_{2month} = \mu^2$ and $\mu_{all} = \mu^3$ and $\mu \in [0, 1]$. We set μ as e^{-1} in the same manner as [6,23], for our experiment.

Interest Propagation: (1) Enabled, or (2) Disabled. In [23], the authors also proposed different propagation strategies exploiting different types of background knowledge from DBpedia. Overall, the propagation strategy extending *primitive interests* with categories (Fig. 2(a)) and entities connected via different properties (Fig. 2(b)) in DBpedia, provided the best performance compared to other state-of-art propagation strategies.

As previous studies [19,22] showed that a discounting strategy is required for the extended concepts based on primitive interests, the authors [23] applied a discounting strategy from [22] for the extended categories as follows:

$$CategoryDiscount = \frac{1}{\alpha} \times \frac{1}{\log(SP)} \times \frac{1}{\log(SC)} \tag{2}$$

where: *SP = Set of Pages belonging to the Category, SC = Set of Sub-Categories.* We set the parameter $\alpha = 2$ as in [23]. Thus, an extended category is discounted



(a) Category-based (b) Property-based

Fig. 2. Three core strategies using DBpedia for extending user interests

heavily if it is a general one, i.e., the category has a great number of pages or subcategories. In addition, the parameter α denotes the discount of the propagated interests from primitive interests. Regarding the property-based extension strategy (Fig. 2(b)), extended entities via different properties are discounted based on the occurrence frequency of a specific property in DBpedia [23]:

$$PropertyDiscount = \frac{1}{\alpha} \times \frac{1}{\log(P)} \tag{3}$$

where: P = *the number of occurrences of a property in the whole DBpedia graph.* The intuition behind *PropertyDiscount* is that entities extended via a property appearing rarely in the DBpedia graph should be given a higher weight than ones extended via a property appearing frequently.

4 Experiment Setup

In the following section, we describe the Twitter dataset used in our experiment (Sect. 4.1), and the evaluation methodology (Sect. 4.2). Subsequently, we present the results using 16 different user modeling strategies in the context of link recommendations on Twitter (Sect. 4.3).

4.1 Twitter Dataset

The dataset used in this experiment is a Twitter dataset from [22], which includes over 340,000 tweets from 480 *active* users on Twitter. An active user denotes that the user published at least 100 Twitter posts [11,15,22]. Table 4 shows the basic statistics about the dataset.

Dataset for link recommendations. In the same way as [23], we further selected users who shared at least one link in their tweets during the previous two weeks, leaving 322 users for our experiment to run upon. We limit our consideration to links having at least four concepts to filter out non-topical links that were automatically generated by third-party applications such as Swarm[14].

[14] https://www.swarmapp.com.

Table 4. Twitter dataset statistics

# of users	480
Total # of tweets	348,554
Average time span of tweets per user (days)	471
Average # of tweets per user	726
Average # of tweets per user per day	7.2

4.2 Evaluation Methodology

We were interested in finding whether combinations of different user modeling dimensions improve the quality of user interest profiles in the context of link recommendations. Therefore, the input to our link recommender system is user interest profiles generated by different user modeling strategies, whereas the output is recommended links (URLs) for users. A lightweight content-based algorithm, like the one used in [5], was applied for recommendations.

Definition 2. *Recommendation Algorithm: given a user profile* P_u *and a set of candidate links* $N = \{P_{i1}, \ldots, P_{in}\}$*, which are represented via profiles using the same vector representation, the recommendation algorithm ranks the candidate items according to their cosine similarity to the user profile.*

We assumed a user was interested in the content of a link if the link was shared by the user in his or her tweets. The ground truth of links was a set of links shared via the user's tweets within the last two weeks, which consists of 3,959 links. Tweets before the last two weeks were used for building user interest profiles. To construct candidate links for recommendations, we further included the links shared by other users but not shared by 322 users in the dataset in addition to the ground truth links from 322 users. The resulting candidate set of links consists of 15,440 distinct links.

The link recommender system measures similarities between a user interest profile and each candidate link, and then provides top-N recommendations based on the similarity scores. We focused on $N = 10$ in our experiment, i.e., the recommendation system would list 10 link recommendations to a user. We measure the quality of recommendations by looking at four different metrics, which were frequently used in the literature [3,5,19,21,23].

- **MRR** The *MRR* (Mean Reciprocal Rank) indicates at which rank the first item *relevant* to the user occurs on average.
- **S@N** The Success at rank N (S@N) stands for the mean probability that a relevant item occurs within the top-N ranked.
- **R@N** The Recall at rank N (R@N) represents the mean probability that *relevant* items are successfully retrieved within the top-N recommendations.
- **P@N** The Precision at rank N (P@N) represents the mean probability that retrieved items within the top-N recommendations are *relevant* to the user.

We set a significance level of alpha = 5 % for all statistical tests. The *bootstrapped paired t-test*[15] was used for testing the significance.

5 Results

In this section, we present the results of experiments using different user modeling strategies in the context of link recommendations. In the following, let *um*(*representation*, *enrichment*, *dynamics*, *semantics*) denote a user modeling strategy where four parameters: *representation*, *enrichment*, *dynamics* and *semantics* represent the four dimensions *Interest Representation*, *Content Enrichment*, *Temporal Dynamics* and *Interest Propagation*, respectively. We use "none" to denote a certain dimension is disabled. For instance, *um*(*concept*, *none*, *none*, *none*) denotes a user modeling strategy using concepts for *Interest Representation* without considering any other dimensions. *um*(*synset* & *concept*, *enrichment*, *none*, *none*) denotes a user modeling strategy using synsets and concepts for *Interest Representation*, and tweets are enriched by the content of embedded links when extracting user interests (i.e., the dimension *Content Enrichment* is enabled).

Table 5 summarizes the recommendation performance using the 16 user modeling strategies in terms of different evaluation metrics. The results are sorted in descending order in terms of MRR. Overall, the best performing strategy is *um*(*synset* & *concept*, *enrichment*, *dynamics*, *none*), which uses DBpedia concepts and WordNet synsets for *Interest Representation*, and considers all other dimensions except *Interest Propagation*. Table 5 shows the importance of (1) *Content Enrichment*, and (2) *Interest Representation* in user modeling. For instance, the strategies enriching tweets with embedded links (1–8 in Table 5) have better performance than the ones without any enrichment (9–16), using the same option for *Interest Representation*. In terms of *Interest Representation* with or without *Content Enrichment*, we observe that using DBpedia concepts with WordNet synsets (1–4 and 9–12) always provides better performance than using concepts alone (5–8 and 13–16). In line with previous work [21], exploiting semantic and lexical knowledge from DBpedia as well as WordNet for *Interest Representation* improves the quality of user modeling.

Table 6 further illustrates statistical differences between the 16 user modeling strategies in terms of MRR. Overall, the results of other evaluation metrics are similar to the MRR and thus omitted for reasons of brevity. The vertical and horizontal dimensions of the table show the comparison between the 16 strategies. As we can see from the table, there are various significant differences between the strategies ($p < .05$, marked in bold font). For example, strategies using concepts and synsets for the dimension *Interest Representation* always significantly outperform strategies using concepts, when other dimensions are kept the same (e.g., 1 and 5). The dimension *Interest Propagation* plays an important role when we use concepts for *Interest Representation* without *Content Enrichment* (13–16). However, when we have a rich interest representation

[15] http://www.sussex.ac.uk/its/pdfs/SPSS_Bootstrapping_22.pdf.

Table 5. Performance of link recommendations using 16 user modeling strategies four different evaluation metrics. The results are sorted in descending order in terms of MRR.

	User Modeling Strategies	MRR	S@10	R@10	P@10
1.	um(synset & concept, enrichment, dynamics, none)	0.3251	0.5062	0.1700	0.1304
2.	um(synset & concept, enrichment, dynamics, propagation)	0.3198	0.4938	0.1654	0.1298
3.	um(synset & concept, enrichment, none, none)	0.3146	0.4876	0.1595	0.1286
4.	um(synset & concept, enrichment, none, propagation)	0.3107	0.4752	0.1534	0.1267
5.	um(concept, enrichment, dynamics, none)	0.2942	0.4193	0.1405	0.1047
6.	um(concept, enrichment, none, none)	0.2886	0.4379	0.1392	0.1062
7.	um(concept, enrichment, dynamics, propagation)	0.2802	0.3975	0.1287	0.0988
8.	um(concept, enrichment, none, propagation)	0.2736	0.4130	0.1332	0.1006
9.	um(synset & concept, none, dynamics, none)	0.2511	0.4255	0.1257	0.0988
10.	um(synset & concept, none, dynamics, propagation)	0.2502	0.4193	0.1259	0.0997
11.	um(synset & concept, none, none, none)	0.2436	0.4068	0.1231	0.0978
12.	um(synset & concept, none, none, propagation)	0.2386	0.3913	0.1179	0.0984
13.	um(concept, none, none, propagation)	0.2083	0.3540	0.0993	0.0820
14.	um(concept, none, dynamics, none)	0.2031	0.3354	0.0927	0.0752
15.	um(concept, none, dynamics, propagation)	0.2024	0.3478	0.0923	0.0795
16.	um(concept, none none, none)	0.1518	0.2609	0.0660	0.0553

(i.e., using concepts and synsets together) or rich content by enrichment, *Interest Propagation* has little effect on the quality of user modeling, i.e., there is no statistical difference between a user modeling strategy with *Interest Propagation* and one without any propagation (1–12). One of the possible reasons might be the rich interest representation, and content is giving sufficient knowledge of user interests. Additionally, the "insufficient quality" of extracted DBpedia entities from tweets using APIs (see the precision in Table 3 in Sect. 3.2), could result in inaccurate interest propagation based on the incorrect entities. This might limit the contribution of propagated interests towards user modeling.

Similar results can be found for temporal dynamics. Although considering *Temporal Dynamics* increases the performance significantly when we use concepts for *Interest Representation* without *Content Enrichment* (13–16), there is no significant difference between strategies with a rich interest representation and rich content (1–12). Nevertheless, we observe that in all of the cases using concepts and synsets for *Interest Representation*, considering the dimension *Temporal Dynamics* provides the best performance (see 1, 9 in Table 5).

To sum up, the two dimensions *Interest Representation* and *Content Enrichment* play significant roles for user modeling, followed by *Temporal Dynamics*. Although the contribution of content enrichment via embedded links might depend on the percentage of embedded links, it is an important and valuable source for enrichment as a large number of tweets are posted with links[16]. The results also show that the *Interest Propagation* dimension had little effect on user modeling when considering different dimensions together, which is different from previous studies considering one or two dimensions [2, 19, 22, 23].

[16] 70 % of one million tweets from U.S. West Coast included links. http://tnw.to/s3R2i.

Table 6. Results of p-values over the 16 user modeling strategies in terms of link recommendations on Twitter (marked in bold font if $p < .05$). Strategies are sorted by MRR results as shown in Table 5.

		2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1.	um(synset & concept, enrichment, dynamics, none)	.14	.17	.11	**.01**	**.02**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**
2.	um(synset & concept, enrichment, dynamics, propagation)		.35	.21	**.04**	**.04**	**.01**	**.01**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**
3.	um(synset & concept, enrichment, none, none)			.24	.10	.05	**.03**	**.01**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**
4.	um(synset & concept, enrichment, none, propagation)				.18	.10	**.03**	**.02**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**	**.00**
5.	um(concept, enrichment, dynamics, none)					.31	.05	**.03**	**.02**	**.02**	**.01**	**.01**	**.00**	**.00**	**.00**	**.00**
6.	um(concept, enrichment, none, none)						.26	.05	**.03**	**.02**	**.01**	**.01**	**.00**	**.00**	**.00**	**.00**
7.	um(concept, enrichment, dynamics, propagation)							.26	.10	.08	.05	**.03**	**.00**	**.00**	**.00**	**.00**
8.	um(concept, enrichment, none, propagation)								.13	.13	.07	**.04**	**.00**	**.00**	**.00**	**.00**
9.	um(synset & concept, none, dynamics, none)									.42	.20	.08	**.01**	**.00**	**.00**	**.00**
10.	um(synset & concept, none, dynamics, propagation)										.22	.08	**.01**	**.01**	**.00**	**.00**
11.	um(synset & concept, none, none, none)											.15	**.02**	**.01**	**.01**	**.00**
12.	um(synset & concept, none, none, propagation)												**.04**	**.03**	**.02**	**.00**
13.	um(concept, none, none, propagation)													.32	.27	**.00**
14.	um(concept, none, dynamics, none)														.46	**.00**
15.	um(concept, none, dynamics, propagation)															**.00**
16.	um(concept, none, none, none)															

6 Conclusions

In this paper, we investigated different combinations of four dimensions of user modeling on Twitter: (1) *Interest Representation*, (2) *Content Enrichment*, (3) *Temporal Dynamics of user interests*, and (4) *Interest Propagation*, which have not been studied together. As a result, we end up with 16 different user modeling strategies with all possible combinations (see Table 2). These strategies were evaluated in the context of link recommendations on Twitter. The best-performing strategy is *um*(*synset* & *concept*, *enrichment*, *dynamics*, *none*), which uses DBpedia concepts and WordNet synsets for *Interest Representation* considering *Temporal Dynamics*, with *Content Enrichment*. The results also indicate that *Interest Representation* and *Content Enrichment* are the most important dimensions compared to other dimensions. In future research, we would like to further investigate how different percentages of links in tweets affect the quality of user modeling.

Acknowledgments. This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight Centre for Data Analytics).

References

1. Abdel-Hafez, A., Xu, Y.: A survey of user modelling in social media websites. Comput. Inf. Sci. **6**(4), 59 (2013)
2. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing temporal dynamics in Twitter profiles for personalized recommendations in the social web. In: Proceedings of 3rd International Web Science Conference, p. 2. ACM (2011)
3. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing user modeling on Twitter for personalized news recommendations. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 1–12. Springer, Heidelberg (2011)
4. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Semantic enrichment of Twitter posts for user profile construction on the social web. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 375–389. Springer, Heidelberg (2011)
5. Abel, F., Hauff, C., Houben, G.-J., Tao, K.: Leveraging user modeling on the social web with linked data. In: Brambilla, M., Tokuda, T., Tolksdorf, R. (eds.) ICWE 2012. LNCS, vol. 7387, pp. 378–385. Springer, Heidelberg (2012)
6. Ahmed, A., Low, Y., Aly, M., Josifovski, V., Smola, A.J.: Scalable distributed inference of dynamic user interests for behavioral targeting. In: Proceedings of 17th International Conference on Knowledge Discovery and Data Mining, pp. 114–122. ACM (2011)
7. Budak, C., Kannan, A., Agrawal, R., Pedersen, J.: Inferring user interests from microblogs. Technical report (2014)
8. Degemmis, M., Lops, P., Semeraro, G.: A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation. User Model. User-Adap. Inter. **17**(3), 217–255 (2007)
9. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 54–89. Springer, Heidelberg (2007). doi:10.1007/978-3-540-72079-9_2
10. Harvey, M., Crestani, F., Carman, M.J.: Building user profiles from topic models for personalised search. In: Proceedings of 22nd ACM International Conference on Information and Knowledge Management, pp. 2309–2314 (2013)

11. Jain, P., Kumaraguru, P., Joshi, A.: @i seek 'fb.me': identifying users across multiple online social networks. In: Proceedings of 22nd International Conference on World Wide Web companion, pp. 1259–1268. ACM (2013)
12. Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A.: User interests identification on Twitter using a hierarchical knowledge base. In: Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8465, pp. 99–113. Springer, Heidelberg (2014)
13. Kinsella, S., Wang, M., Breslin, J.G., Hayes, C.: Improving categorisation in social media using hyperlinks to structured data sources. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 390–404. Springer, Heidelberg (2011)
14. Locke, B.W.: Named entity recognition: adapting to microblogging. Ph.D. thesis (2009)
15. Lu, C., Lam, W., Zhang, Y.: Twitter user modeling and tweets recommendation based on Wikipedia concept graph. In: Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence (2012)
16. Michelson, M., Macskassy, S.A.: Discovering users' topics of interest on Twitter: a first look. In: Proceedings of Fourth Workshop on Analytics for Noisy Unstructured Text Data, pp. 73–80. ACM (2010)
17. Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: inferring user profiles in online social networks. In: Proceedings of Third ACM International Conference on Web Search and Data Mining, pp. 251–260. ACM (2010)
18. Nishioka, C., Scherp, A.: Profiling vs. time vs. content: what does matter for top-k publication recommendation based on Twitter profiles? In: Proceedings of 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, JCDL 2016, pp. 171–180. ACM, New York (2016)
19. Orlandi, F., Breslin, J., Passant, A.: Aggregated, interoperable and multi-domain user profiles for the social web. In: Proceedings of 8th International Conference on Semantic Systems, pp. 41–48. ACM (2012)
20. Piao, G.: Towards comprehensive user modeling on the social web for personalized link recommendations. In: Proceedings of 2016 Conference on User Modeling Adaptation and Personalization, UMAP 2016, pp. 333–336. ACM (2016)
21. Piao, G.: User modeling on Twitter with WordNet Synsets and DBpedia concepts for personalized recommendations. In: The 25th ACM International Conference on Information and Knowledge Management. ACM (2016)
22. Piao, G., Breslin, J.G.: Analyzing aggregated semantics-enabled user modeling on Google+ and Twitter for personalized link recommendations. In: User Modeling, Adaptation, and Personalization, pp. 105–109. ACM (2016)
23. Piao, G., Breslin, J.G.: Exploring dynamics and semantics of user interests for user modeling on Twitter for link recommendations. In: 12th International Conference on Semantic Systems, pp. 81–88. ACM (2016)
24. Zarrinkalam, F., Kahani, M.: Semantics-enabled user interest detection from Twitter. In: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (2015)