

Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web

Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao
Web Information Systems, TU Delft
PO Box 5031, 2600 GA Delft, the Netherlands
{f.abel,q.gao,g.j.p.m.houben,k.tao}@tudelft.nl

ABSTRACT

Social Web describes a new culture of participation on the Web where more and more people actively participate in publishing and organizing Web content. As part of this culture, people leave a variety of traces when interacting with (other people via) Social Web systems. In this paper, we investigate user modeling strategies for inferring personal interest profiles from Social Web interactions. In particular, we analyze individual micro-blogging activities on Twitter. We compare different strategies for creating user profiles based on the Twitter messages a user has published and study how these profiles change over time. Moreover, we evaluate the quality of the user modeling strategies in the context of personalized recommender systems and show that those strategies which consider the temporal dynamics of the individual profiles allow for the best performance.

1. INTRODUCTION

With more than 190 million users and more than 65 million postings per day, Twitter is clearly one of the most popular services on the Social Web¹. To understand this Social Web and the way in which humans are part of it, an analysis of Twitter is an effective instrument for scientists. Twitter poses a simple question to its users, “what’s happening?”, and restricts the answer to this question to 140 characters. People therefore publish short messages (tweets) about their everyday activities on Twitter. Other users follow selected information streams and can react to Twitter messages, e.g. by re-tweeting or posting a reply. Lately, researchers study the network structures that evolve from those user interactions on Twitter [4, 13] and investigate how information propagates through the Twitter network [9, 10]. Yet, little research has been done on understanding the semantics of individual Twitter activities and inferring user interests from these activities, user interests that can for example be used in modeling the users as basis for personalized recom-

mendations in (other) applications in the Web. Given the shortness of tweets, making sense of individual tweets and exploiting tweets for user modeling are non-trivial problems. In this paper, we describe the research questions that underlie our studies and evaluations. We report on the findings in a set of studies related to personalized news recommendations where we in particular consider the temporal dynamic effects.

Our ambition is to understand how people behave in the Social Web. In previous work, we analyzed the nature of tagging activities that people perform on Social Web systems like Flickr, Delicious and StumbleUpon and studied the impact of cross-system user modeling on personalization [3].

In this paper, we study characteristics of Twitter-based profiles and investigate how one can leverage Twitter activities for personalization. For this purpose, we developed a library for aggregating and enriching the semantics of individual Twitter activities [2]. Given this work, we conduct an in-depth analysis on a large Twitter dataset of more than 30 million tweets that were published by more than 20,000 users in a period of four months². We study user modeling on Twitter and answer research questions that concern the temporal evolution of individual user profiles inferred from Twitter activities:

- How can we infer personal interests from individual Twitter interactions and to what extent do personal interests vary over time?
- How are personal interests and user concerns influenced by public trends? Do inferred interests profiles allow for predicting which trends will be adopted by a user?
- Can we exploit Twitter-based interest profiles to personalize the users’ Social Web experiences? How do temporal dynamics impact the accuracy of personalization?

To answer the above research questions, we develop a framework for enriching and contextualizing Twitter messages. Our framework allows us to identify entities (e.g. Muammar Gaddafi (person), Apple (company)) and topics (e.g.

¹<http://techcrunch.com/2010/06/08/twitter-190-million-users/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebSci ’11, June 14-17, 2011, Koblenz, Germany.
Copyright 2011 ACM.

²We make a subset of our dataset publicly available at <http://wis.ewi.tudelft.nl/websci11/>

politics, sports) that are mentioned by the users in the individual tweets [2]. Moreover, we are able to relate tweets to news events (e.g. protests in Egypt) and external Web resources to further contextualize the users' Twitter activities. The semantically enriched Twitter interactions form the basis for a variety of user modeling strategies that we analyze in this paper to understand users' individual Twitter interactions over time.

The main contributions of our work can be summarized as follows.

- We develop a framework for modeling the semantics of individual Twitter interactions. Our framework features a variety of user modeling strategies that enrich the semantics of individual Twitter messages, capture personal user interests over time and relate personal user interests with global trends.
- We conduct a large scale analysis of Twitter-based user modeling and analyze different design dimensions of the user modeling strategies such as weighting schemes and semantic enrichment strategies in detail. We particularly study the temporal dynamics of Twitter-based user interest profiles and analyze the relations between personal interests and the adoption of trending topics.
- We prove the success of our strategies in the context of personalized news recommendations on the Social Web. We evaluate the impact of the different design choices on personalization quality in the context of recommending Web sites and show that the discovered temporal features can successfully be exploited to improve the accuracy of personalization.

The rest of the paper is organized as follows: based on an analysis of the temporal characteristics of user interests into trending topics presented in the next section, we will introduce the core model of our user modeling framework in Section 3. Our framework allows for the creation of user modeling strategies that consider temporal dynamics of user behavior on Twitter. In Section 4 we study how the specifics of the different user modeling strategies impact personalization before we conclude in Section 5.

2. EVOLUTION OF USER INTERESTS IN TRENDING TOPICS

To better understand the temporal dynamics of the interests and concerns individual users express on Twitter by posting messages, we monitored – starting from November 15th 2010 – more than 20,000 users over a period of more than four months and overall collected more than 30 million tweets. In this section, we analyze how the interests of users into a topic discussed on Twitter change over time. Therefore, we start with a concrete example topic, the Egyptian revolution³, which started on January 25th, 2011. In this analysis, we aim to clarify on the following research questions.

1. How can a topic on Twitter be represented?

³http://en.wikipedia.org/wiki/Timeline_of_the_2011_Egyptian_revolution

2. How do topics evolve over time and how does this affect the representation of a topic?
3. How do the interests of individual users into a topic change over time?

We base our analysis on previous work, in which we studied how to extract trends from Twitter [7]. For representing a trending topic, it is often not sufficient to represent it just via a single concept such as a hashtag (words starting with “#”). For example, regarding the Egyptian revolution a hashtag like “#egypt” could be considered as a representative concept to describe this topic. However, (i) not all tweets that contain the hashtag “egypt” refer to the revolution in Egypt and (ii) there exist tweets that refer to the revolution but do not mention the hashtag “#egypt”. Instead, other terms that refer to entities such as *Mubarak* (person) or *Cairo* (location) may be used. Therefore, we propose to model a topic on Twitter as a set of weighted concepts where a concept may refer to an arbitrary entity and where the weight indicates how important the concept is for the topic:

DEFINITION 1 (TOPIC). *A topic is a set of weighted concepts where a concept c may be represented via a named entity or hashtag.*

$$\text{topic}(\text{time}, T_{\text{tweets}}) = \{(c, w(c, \text{time}, T_{\text{tweets}})) | c \in C_H \cup C_E\} \quad (1)$$

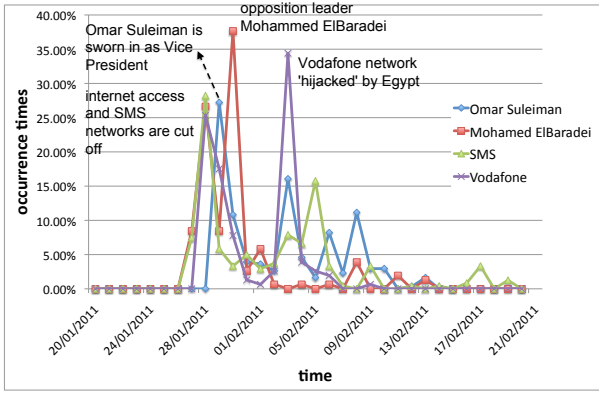
Here, $w(c, \text{time}, T_{\text{tweets}})$ is a function that computes the weight associated to the concept c for the topic based on messages T_{tweets} that are (possibly) related to the topic and based on a given timestamp. C_H and C_E denote the set of hashtags and entities respectively.

The above model for creating the representation of a topic expects a timestamp as input because concepts that relate to a certain topic may change over time. On the one hand, the importance of concepts could vary at different points in time and, on the other hand, new concepts could arise while other concepts that were once representing the topic could entirely become useless to describe the topic. Hence, the representation of a topic depends on the time when the profile is demanded.

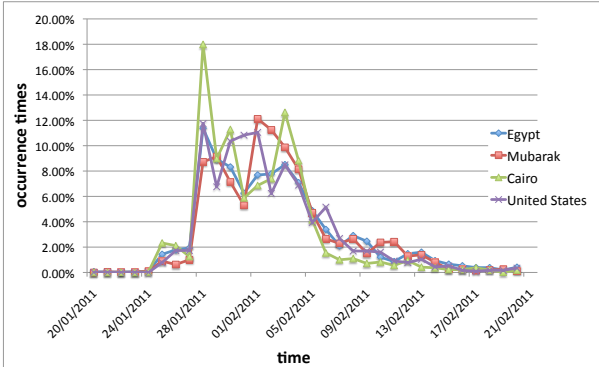
2.1 Evolution of Topics over Time

To analyze how the topic *Egyptian revolution* evolves over time, we selected popular entities on every day of our observation period based on their co-occurrence frequency with hashtags such as “#jan25” or “#tahrir” which we could almost unambiguously relate to the topic.

Figure 1 illustrates how the occurrence frequency of entities, which are related to the Egyptian revolution, changes over time. Some entities like *Cairo* and *Mubarak* are popular for this topic over a long period in time (see Figure 1(b)), which means that Twitter users continuously refer to these entities when publishing tweets about the topic. The occurrence frequencies of these entities quickly reach their peaks three days after the beginning of the Egyptian revolution, which started on January 25th, and then decrease rather slowly over the next two weeks.



(a) Entities with short lifespan for the topic.



(b) Entities with long lifespan for the topic.

Figure 1: Relative occurrence frequencies of entities related to the Egyptian revolution.

In contrast, the entities shown in Figure 1(a) show burst-like spikes and seem to be relevant for the topic only for a short period in time. For example, many messages that were posted on January 28th were related to the entity *SMS* and referred to the shutdown of the Internet access and short messaging services in Egypt that happened on January 26th, such as the following tweet:

“Again, latest Egypt updates: internet shut down, SMS and Blackberry down, plainclothes police setting cars on fire”

Therefore the entity *SMS* became very popular on that day. Similarly, *Omar Suleiman* was mentioned in many messages on January 30th as he was sworn in as vice president on January 29th which resulted in further protests as reported in the following message:

“Al Jazeera breaking: Protesters loudly condemn the appointment of Omar Suleiman as Vice President”

Similarly, the leader of the opposition *Mohamed ElBaradei* became popular for the topic on January 30th as well. Moreover, the peak for *Vodafone* on February 3rd is much likely related to the news that the mobile phone company an-

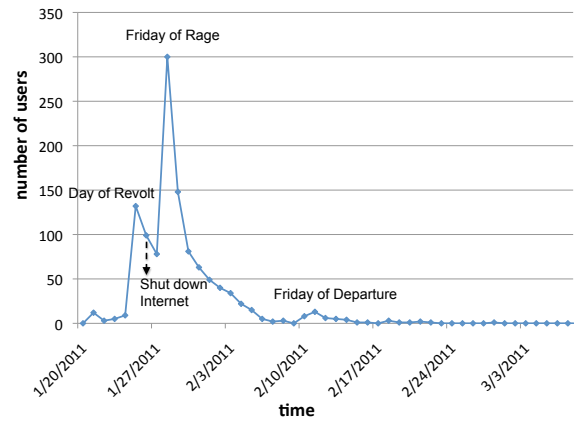


Figure 2: User adoption: number of new users per day who become interested in the Egyptian revolution.

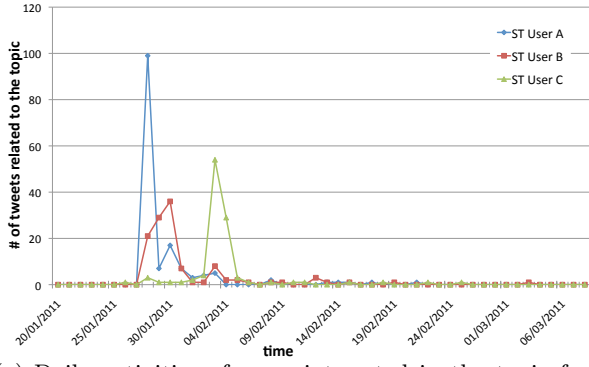
nounced that the Egyptian authorities had hijacked Vodafone’s network.

Our analysis presented in Figure 1 thus demonstrates that the importance of entities for a given topic changes over time. While there are some entities that are continuously good representatives for a topic (e.g. *Mubarak*), there are other entities (e.g. *SMS*) which characterize a topic only for a short period in time. When creating the representation of a topic it is thus reasonable to consider multiple concepts (e.g. entities and hashtags) and to compute the importance of each concept as a function of the time when the topic representation is requested.

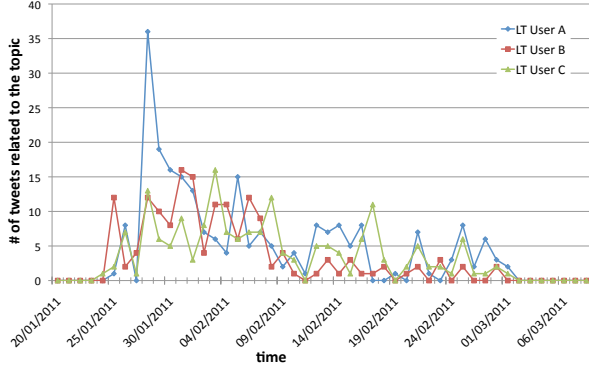
2.2 Evolution of User Interests into Topics

Having seen how a topic is discussed within a community of users and how the representation of a topic emerges and changes over time, we now analyze how the interests of individual users into a topic evolves over time. We therefore selected a subset of 1619 Twitter users. In particular, those users for which we monitored at least 20 Twitter messages in total and observed at least 10 Twitter messages during the time of the Egyptian revolution but not necessarily 10 messages that are related to the incident in Egypt. In fact, we discovered that 70% of the sample users showed interest into the Egyptian revolution, i.e. 70% of the users (re-)tweeted a message that was mentioning a concept of the corresponding topic representation. While these users were interested in the topic, the individual behavior showed interesting specifics. For example, not all the users started tweeting about the event from the very beginning (January 25th). Figure 2 shows for each day the number of users who published their first tweet about the Egyptian revolution and therefore showed for the first time that they are – to some extent – concerned with the topic.

As shown in Figure 2, most people do not join the discussion or dissemination of the event immediately after it happens. While the small amplitudes before January 25th can be considered as noise and seem to be caused by the modeling of the topic, on the day of the first wave of protest in Egypt, the “Day of Revolt”, slightly less than 150 of our sample users joined the discussion on the topic. After the Eyp-



(a) Daily activities of users interested in the topic for a short time period (*short-term adopters*).



(b) Daily activities of users interested in the topic for a long time period (*long-term adopters*).

Figure 3: Daily activities of users who are interested in the Egyptian revolution.

tian regime shut down the Internet on January 26th, about 300 users became interested into the protests on the “Friday of Rage”, January 28th, and another 150 users took for the first time part in the Twitter discussions on the following day.

Having seen *when* individual users become for the first time interested in a topic, we were also interested for *how long* those users were interested in the topic. Figure 3 shows the amount of Twitter messages that selected users were posting on different days. The users whose tweeting activities on the topic of the Egyptian revolution are displayed in Figure 3(a) can be characterized as *short-term adopters* as they published tweets about the event for less than one week. It is interesting to see that the amount of messages these users posted about the topic is fairly high. For example, *ST User A*, who adopted the topic two days after the beginning of the revolt, published almost 100 tweets about the revolution on a single day. Nevertheless, she quickly became disinterested. The interests of these three example users thus seem to change quickly. Hence, user modeling strategies that aim for capturing users’ interests into topics have to adapt quickly as well.

Figure 3(b) displays the Twitter activities of three other users who were concerned with the Egyptian revolts for a long time period of more than one month and can therefore

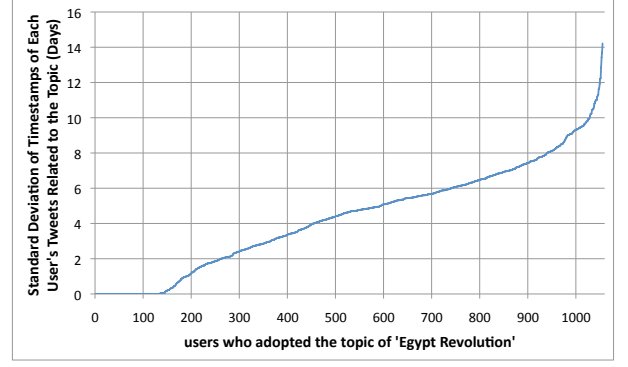


Figure 4: Standard Deviation of Timestamps of Related Tweets Posted by Each User

be considered as *long-term adopters*. All the three long-term adopters became interested into the topic at the very beginning of the revolt and can thus also be described as *early adopters*. In contrast, the short-term adopters characterized in Figure 3(a) are not among the first users who publish about the incidents in Egypt. In fact, for the Egyptian revolution it seems that there is a correlation between the time *when* a user adopts a topic and the duration during which the user is interested into the topic, i.e. early adopters overlap stronger with long-term adopters than with short-term adopters. Furthermore, the Twitter behavior of the short-term adopters regarding the Egyptian revolution is apparently more influenced by public trends than the behavior of the long-term adopters. For example, as depicted in Figure 3(a), *ST User A*, *B* and *C* show a peak after the riot on February 2nd that was entitled the “*Battle of the Camel*” and which was heavily discussed in social and mainstream news media. In contrast, the peaks of the long-term adopters, shown in Figure 3(b), happen much more frequently and also occur on days on which were not packed with epic events.

Figure 4 overviews the sample users who were interested in the Egyptian revolution with respect to the duration the different users expressed their interest into the topic on Twitter. In particular, it shows for each user the standard deviation of the timestamps of tweets that were related to the topic as similarly proposed by Huang et al. [8] who measure the temporal stability of hashtags. For Figure 4, we apply standard deviation as follows.

$$\sigma(\text{topic}, \text{user}) = \sqrt{\frac{\sum_{k=1}^N (\text{time}(\text{tweet}_k) - \overline{\text{time}})^2}{N - 1}} \quad (2)$$

Here, $\text{time}(\text{tweet}_k)$ is the timestamp of the k -th tweet published by the given *user* that refers to the given *topic*, $\overline{\text{time}}$ is the average timestamp of the user’s tweets that relate to the topic and N is the overall number of tweets in which the user refers to the topic.

Figure 4 shows that for nearly 150 users the $\sigma(\text{topic}, \text{user})$ is zero which means that those users just published one tweet that we could relate to the happenings in Egypt. Overall, for more than 75% of the users, the standard deviation of times-

tamps which specify when they published about the topic is less than one week. The fraction of long-term adopters for whom $\sigma(\text{topic}, \text{user})$ is higher than ten days is with less than 2.5% rather low.

2.3 Findings

In summary, we can thus answer the research questions raised at the beginning of this section as follows.

1. Topics that are discussed on Twitter can be represented via the concepts that are referenced from the tweets that relate to the topic. Those concepts can be arbitrary entities such as persons, organizations or locations as well as cryptic hashtags like “#jan25”. As different concepts may be of different relevance for a topic, it is desirable to weigh the concepts according to their importance for the topic.
2. Topics change over time: different concepts are of different importance for a given topic. For example, concepts such as *SMS* or *Vodafone* became important for the Egyptian revolution only for a short time when the government of Egypt shut down the Internet and took over the telecommunication network of Vodafone. Due to this event-like nature of a Twitter topic, it is helpful to compute the weight of a concept for a topic as a function of time.
3. The interests of individual users into a topic evolve differently over time in the context of the Egyptian revolution. Most users, who were interested in the topic, adopted the topic within a few days. Hence, the speed in which people adopt a topic on Twitter seems to be rather fast (cf. [9, 12]). However, the fraction of early adopters who become interested in an event on the day the event happens is small. Moreover, the duration during which users are interested in an event-like topic differs clearly among the different users. In fact, we identified long-term adopters who are interested in a Twitter topic over a long period in time and short-term adopters who are concerned with a topic only for a short period in time and are rather driven by current trends.

3. USER MODELING WITH TEMPORAL DYNAMICS

Given the findings presented in the previous section, we now introduce a lightweight user modeling framework that allows for the creation of strategies that infer user interests from the Twitter activities of a user and allow for capturing temporal dynamics in these profiles. We implemented our approach as extension to the Twitter-based user modeling framework introduced in [1] and make our strategies also available via Web services⁴.

Our user modeling strategies aggregate and monitor Twitter messages of an individual user and process each tweet by means of a semantic enrichment pipeline that extracts hashtags and named entities (e.g. persons, locations or organizations) from a given tweet. As Twitter messages

are limited to 140 characters it may become difficult to extract meaningful concepts from the tweets. For example, given a tweet such as “President’s son and family flee: <http://fb.me/J6SmQF7q>” it is difficult to understand to which president, son and family the user refers to. However, the semantics of the message can be interpreted when following the link posted in the tweet. In this paper, we thus include a semantic enrichment component into the Twitter user modeling process that follows hyperlinks, extracts the main content of the linked Web pages and identifies entities mentioned in those pages. This allows us to represent a tweet via both concepts which are extracted from the tweet and concepts that are extracted from Web sites that are referenced from the tweet. Based on the semantically enriched Twitter messages of a user, we create strategies that infer user interest profiles. In this paper, we represent those profiles in the same way as we represent topics.

DEFINITION 2 (USER PROFILE). *The profile of a user u is a set of weighted concepts where a concept c may be represented via a named entity or hashtag.*

$$P(u, \text{time}) = \{(c, w(c, \text{time}, T_{\text{tweets}, u}) | c \in C_H \cup C_E\} \quad (3)$$

Here, $w(c, u, \text{time})$ is a function that computes the weight associated with the concept c for the given user u based on messages $T_{\text{tweets}, u}$ published by u and based on the given timestamp. C_H and C_E denote the set of hashtags and entities respectively.

With $\vec{p}(u, \text{time})$ we refer to $P(u, \text{time})$ in its vector space model representation, where the value of the i -th dimension refers to $w(c_i, \text{time}, T_{\text{tweets}, u})$. A straightforward approach for computing the weight is to determine the occurrence frequency of the concept c in the set of tweets published by the user u . We compare this baseline strategy that ignores the time input with a time-sensitive variant which dampens the occurrence frequency according to the temporal distance between the concept occurrence time and the given timestamp.

$$w(c, \text{time}, T_{\text{tweets}, u}) = \sum_{t \in T_{\text{tweets}, u, c}} \left(1 - \frac{|\text{time} - \text{time}(t)|}{\text{max}_{\text{time}} - \text{min}_{\text{time}}}\right)^d \quad (4)$$

In Equation 4, $T_{\text{tweets}, u, c}$ denotes the set of tweets that have been published by u and refer to the concept c . $\text{time}(t)$ returns the timestamp of a given tweet t and max_{time} and min_{time} denote the highest (youngest) and lowest (oldest) timestamp of a tweet in $T_{\text{tweets}, u, c}$, for example: $\text{max}_{\text{time}} = \max(\{\text{time}(t) | t \in T_{\text{tweets}, u, c}\})$. The parameter d is used to adjust the influence of the temporal distance. The higher d is set, the higher the penalty of concepts that occur with a high distance to the input time as the corresponding scores will be lower than for those concepts for which $|\text{time} - \text{time}(t)|$ is smaller. In the subsequent sections we set $d = 4$. Furthermore, we normalize the weights of a profile $P(u, \text{time})$ so that the sum of weights in a profile is equal to 1.

Our hypothesis is that the time-sensitive strategy characterizes the actual demands and concerns of a user better than the non-time-sensitive baseline strategy.

4. TIME-SENSITIVE USER MODELING FOR PERSONALIZED RECOMMENDATIONS

⁴<http://wis.ewi.tudelft.nl/tums/>

To investigate the above hypothesis, we deploy the user modeling strategies in a personalized recommender system. The recommender provides Web site recommendations to a user based on her user profile. We thus apply the Twitter-based user modeling strategies to personalize the Social Web experience of the users and point them to Web sites which are according to their profiles of interest in their current temporal context. We then study the following research questions.

1. How do semantic enrichment and (time-sensitive) weighting functions of the user modeling framework influence the performance of the recommender system?
2. Are there any correlations between characteristic patterns in the generated Twitter profiles and the gained recommendation quality? For example, how does the recommendation quality differ between users who have a tendency to be short-term or long-term adopters on a given topic?

4.1 Evaluation Methodology

We examine the user modeling strategies in the context of a recommender system that we developed for providing personalized Web site recommendations to the user. In particular those fresh Web sites that are referenced in Twitter messages (cf. [5, 6]). Recommending Web sites, which are posted on Twitter, is a non-trivial task as URLs, which are going to be recommended, often refer to news articles or other types of fresh, news-like content [9]. This makes it difficult to apply collaborative filtering methods, but rather calls for content-based or hybrid approaches [11]. Our main goal is to analyze and compare the applicability of the different user modeling strategies in the context of the recommender system. We particularly analyze how the time-sensitive user modeling strategy, introduced in Section 3, influences personalization and performs in comparison to non-time-sensitive variants. We do not aim to optimize recommendation quality, but are interested in comparing the quality achieved by the same recommendation algorithm when inputting different types of user profiles. Therefore we apply a lightweight content-based algorithm that recommends items according to their cosine similarity with a given user profile. We thus cast the recommendation problem into a search and ranking problem where the given user profile, which is constructed by a specific user modeling strategy, is interpreted as query.

DEFINITION 3 (RECOMMENDATION ALGORITHM). *Given a user profile vector $\vec{p}(u)$ and a set of candidate Web resources (URLs) $R = \{\vec{p}(r_1), \dots, \vec{p}(r_n)\}$, which are represented via profiles using the same vector representation that is used for a given user profile $\vec{p}(u)$, the recommendation algorithm ranks the candidate items according to their cosine similarity to $\vec{p}(u)$.*

$$\text{sim}_{\text{cosine}}(\vec{p}(u), \vec{p}(r_i)) = \frac{\vec{p}(u) \cdot \vec{p}(r_i)}{\|\vec{p}(u)\| \cdot \|\vec{p}(r_i)\|} \quad (5)$$

Given the Twitter dataset which contains more than 30 million tweets and more than 1.3 million distinct Web sites that are linked from the tweets, we compute personalized recommendations for each user of our sample (cf. Section 2) on

each day of our recommendation period which is given by the last ten days of January (Jan 20th - Jan 31st). Hence, our recommendation period overlaps with the beginning of the Egyptian revolution. However, the Web sites that are recommended to the users in this period may refer to any topic and are not necessarily related to the revolution in Egypt. The ground truth of URLs, which we consider as *relevant* for a specific user u on a particular day, is given by those Twitter messages which link to the corresponding Web site and which have been re-tweeted by u on that day. Following this evaluation strategy, we identified, on average, 24.5 relevant URLs for each of the 1619 sample users per day. The candidate set of URLs, which were published on a recommendation day, contained, on average, 24549 items.

Given the ground truth and candidate sets, we applied the different user modeling strategies together with the above algorithm (see Definition 3) and set of candidate items to compute fresh, personalized Web site recommendations for each user on each day. The user modeling strategies were only allowed to exploit tweets published before the start of the recommendation period. The quality of the recommendations was measured by means of $S@k$ (Success at rank k), which stands for the mean probability that a relevant item occurs within the top k of the ranking, and MRR (Mean Reciprocal Rank), which indicates at which rank the first item relevant to the user occurs on average. For Success@ k , we will focus on $S@10$ as our recommendation system will list 10 Web site recommendations to a user.

4.2 Results

Figure 5 summarizes the result of our recommendation experiment. In Figure 5(a), we first analyze the impact of the semantic enrichment provided by our user modeling framework. We observe that the recommendation quality is positively influenced by the enrichment component that follows the links in Twitter messages to also extract named entities from those Web pages. While the performance regarding MRR increases just slightly, $S@10$ improves by more than 15%. For the entity-based user modeling strategy, we thus apply the semantic enrichment method that exploits the links posted in Twitter messages also for the subsequent recommendation experiments.

Figure 5(b) shows the performance of the entity-based and hashtag-based user modeling strategies and illustrates how the time-dependent weighting function (cf. Equation 4) influences the personalization quality. Regarding $S@10$, the entity-based user modeling strategy performs slightly better than the hashtag-based method (improvement: 5%). However, there is no significant difference in performance between entity-based and hashtag-based user modeling strategy. In contrast, the time-dependent weighting function increases the recommendation performance clearly. For the hashtag-based user modeling strategy, weighting the occurrence frequency according to the time for which a profile is demanded (*hashtag (time)*) improves the recommendation quality over the baseline strategy (*hashtag*) by 10.4% and 12% regarding $S@10$ and MRR respectively. We thus find first evidence for our hypothesis that the time-sensitive strategy characterizes the actual demands and concerns of a user better than the non-time-sensitive baseline strategy.

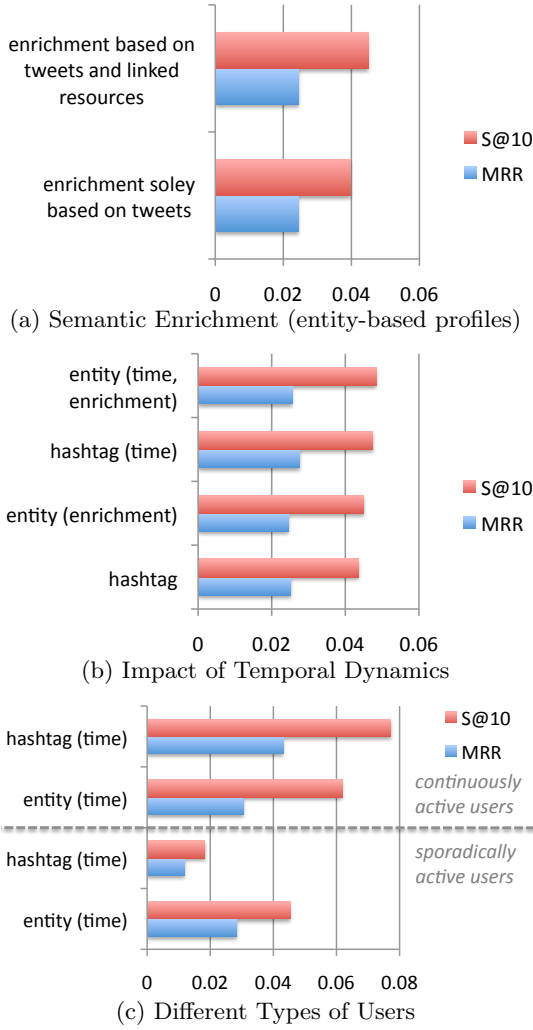


Figure 5: Comparison of user modeling strategies for supporting personalization.

Figure 5(c) illustrates the recommendation performance for different types of users: (i) people who are *continuously active* during our recommendation period and re-tweet at least one Web site on each day of the ten days (i.e. for each day there exist at least one relevant item to be recommended) and (ii) people who are *sporadically active* (on less than five days). As depicted in Figure 5(c), the recommendation performance is better for active users than for the sporadically active users. It is interesting to see that the hashtag-based version performs best for the continuously active users and rather fails for the sporadically active users for which the entity-based user modeling strategy performs best. Hence, it seems that for recommending Web sites on the Social Web, the interests of active users can be represented best via hashtags while the interests of sporadically active users are best modeled via the entity-based strategy.

Figure 6 further relates the recommendation quality with the size of entity-based and hashtag-based profiles. The size of a user's profile is measured by the number of distinct concepts that appear in a profile and is given relatively to the size of the biggest profile. The performance is measured via the mean reciprocal rank (MRR) and is also specified in a

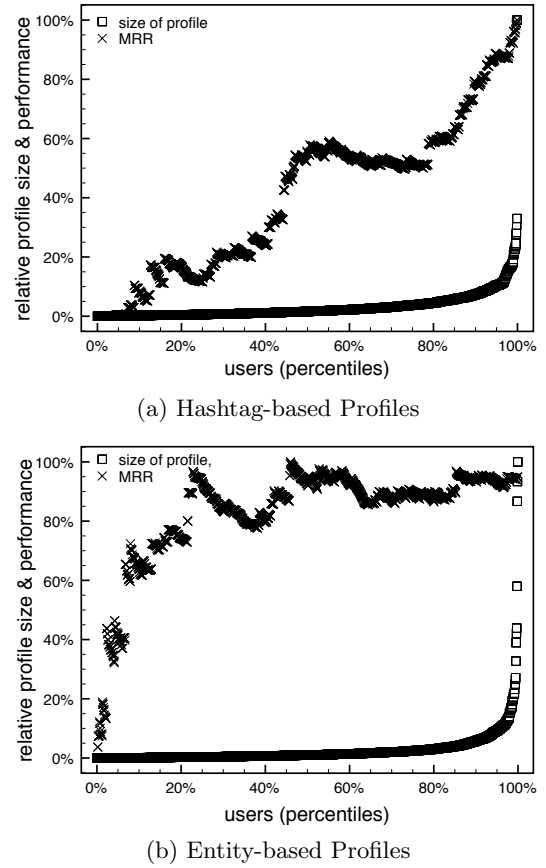


Figure 6: Relation between size of profiles and quality of profiles for supporting personalization.

relative manner. Moreover, the MRR curves show the average performance for the corresponding $x\%$ of the users. For example, for those 20% of the users whose hashtag-based profiles are smaller than the profiles of the other 80% of the users, the recommendation quality is less than 20%. Figure 6(a) can thus be interpreted as follows: the bigger the hashtag-based profiles the better the recommendation.

For entity-based profiles, we observe different behavior, as depicted in Figure 6(b). The quality of recommendations computed based on entity-based profiles does not depend that strongly on the size of the profiles. In fact, it remains fairly stable for varying profile sizes.

4.3 Findings

In this section, we showed how the Twitter-based user modeling strategies can be applied in a recommender system to personalize the users' Social Web experience. The research questions raised at the beginning of this section can be answered as follows.

1. When determining the importance of concepts in a user profile, it is beneficial to weigh the concepts with respect to the point in time for which the profile is demanded. Those concepts which a user has been concerned with recently should be weighted higher than concepts which have not been referenced by the user for a long time. Moreover, we observed that entity-

based user modeling performs best when extracting entities from both the Twitter messages and the Web resources, which are referenced from the corresponding Twitter message.

2. We also discovered remarkable correlations between the characteristics of the different types of user profiles and the resulting recommendation quality. When modeling users based on hashtags, the personalization performance correlates with the size of the hashtag-based profile: the bigger the profile, the better the performance. In contrast, personalization enabled via entity-based user modeling is highly independent from the size of a profile. Furthermore, we observed that for sporadically active users, which tend to be short-term adopters (cf. Section 2), the entity-based user modeling strategies, provided by our framework, perform much better than the hashtag-based strategies.

5. CONCLUSIONS

In this paper, we analyzed the temporal dynamics of user profiles inferred from users' Twitter activities. We presented a user modeling framework that allows for the creation of strategies which extract the semantics of individual Twitter messages and allow for the generation of user interest profiles that specify to which degree a user is interested into a given concept. Given this framework, we first analyzed the characteristics of topics discussed on Twitter and discovered that the representation of a topic changes over time: concepts related to a topic may gain or loose importance. For event-like topics, we identified different groups of users: *long-term adopters* join the discussion early and continuously contribute to the discussion while *short-term adopters* join the discussion later and participate just sporadically being influenced by public trends.

Based on this analysis, we introduced strategies that allow for incorporating those temporal characteristics into user profiles as well. We defined time-sensitive user modeling strategies (hashtag-based and entity-based) and evaluated these strategies in context of a recommender system that provides Web site recommendations on the Social Web. Our results prove the benefits of user modeling strategies that capture the temporal dynamics of a user's Twitter activities and reveal that semantic enrichment is particularly important for users who sporadically participate in the discussions on Twitter.

Acknowledgements. This work is partially sponsored by the EU FP7 project ImREAL⁵.

6. REFERENCES

- [1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing User Modeling on Twitter for Personalized News Recommendations. In *International Conference on User Modeling, Adaptation and Personalization (UMAP)*, Girona, Spain. Springer, 2011.
- [2] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. In *Extended Semantic Web Conference (ESWC)*, Heraklion, Greece. Springer, 2011.
- [3] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction (UMUAI), Special Issue on Personalization in Social Web Systems*, pages 1–42, 2011.
- [4] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In W. W. Cohen and S. Gosling, editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM)*, Washington, DC, USA. The AAAI Press, 2010.
- [5] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the 28th international conference on Human factors in computing systems (CHI)*, pages 1185–1194. ACM, 2010.
- [6] A. Dong, R. Zhang, P. Kolari, J. Bai, F. Diaz, Y. Chang, Z. Zheng, and H. Zha. Time is of the essence: improving recency ranking using twitter data. In *Proceedings of the 19th international conference on World Wide Web (WWW)*, pages 331–340. ACM, 2010.
- [7] Q. Gao, F. Abel, G.-J. Houben, and K. Tao. Interweaving trend and user modeling for personalized news recommendations. Technical report, submitted to the International Conference on Web Intelligence Web (WI), Lyon, France, 2011.
- [8] J. Huang, K. M. Thornton, and E. N. Efthimiadis. Conversational Tagging in Twitter. In M. H. Chignell and E. Toms, editors, *Proceedings of the 21st ACM conference on Hypertext and hypermedia (HT)*, pages 173–178. ACM, 2010.
- [9] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, pages 591–600. ACM, 2010.
- [10] K. Lerman and R. Ghosh. Information contagion: an empirical study of spread of news on digg and twitter social networks. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*, The AAAI Press, 2010.
- [11] J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In C. Rich, Q. Yang, M. Cavazza, and M. X. Zhou, editors, *Proceeding of the 14th international conference on Intelligent user interfaces (IUI)*, pages 31–40. ACM, 2010.
- [12] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web (WWW)*, pages 851–860. ACM, 2010.
- [13] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twiterrank: finding topic-sensitive influential twitterers. In B. D. Davison, T. Suel, N. Craswell, and B. Liu, editors, *Proceedings of the Third International Conference on Web Search and Web Data Mining (WSDM)*, New York, NY, USA, pages 261–270. ACM, 2010.

⁵<http://imreal-project.eu>