Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Twitter-Based User Modeling for News Recommendations*

Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao
Web Information Systems, Delft University of Technology
Mekelweg 4, 2628CD Delft, the Netherlands
{f.abel,q.gao,g.j.p.m.houben,k.tao}@tudelft.nl

Abstract

In this paper, we study user modeling on Twitter. We investigate different strategies for mining user interest profiles from microblogging activities ranging from strategies that analyze the semantic meaning of Twitter messages to strategies that adapt to temporal patterns that can be observed in the microblogging behavior. We evaluate the quality of the user modeling methods in the context of a personalized news recommendation system. Our results reveals that an understanding of the semantic meaning of microposts is key for generating high-quality user profiles.

1 Introduction

People publish short messages on Twitter to share their thoughts and things that happen in their daily life. A plethora of digital traces that people leave in the microblogging sphere provides possibilities for modeling user preferences and delivering personalized services. Some research initiatives show that the exploitation of tweets allows for valuable applications such as early warning system [Sakaki *et al.*, 2010] or discovery of fresh Web sites [Dong *et al.*, 2010]. These applications mainly utilize the wisdom of the crowds as source of information rather than relying on individual tweets.

Learning and modeling the semantics of individual Twitter activities is important for better supporting various applications that aim for personalization. For example, given the huge amount of information disseminated daily on Twitter, user profiling that supports users in ranking sources to follow [Hannon *et al.*, 2010] or selecting content to read [Chen *et al.*, 2010] is becoming crucial. Recently, researchers started to exploit Twitter activities to understand users' preferences and behavioral patterns. Cheng et al. investigate how to infer a user's location based on the content of tweets [Cheng *et al.*, 2010]. Golbeck et al. present a method to measure users' political orientations [Golbeck and Hansen, 2011]. Cha et al. study the dynamics of user influence across topics and time [Cha *et al.*, 2010]. Yet, little research has been done that focuses on understanding the semantics of individual tweets and inferring user interests from these activities. As tweets are limited to 140 characters, making sense of individual tweets and exploiting tweets for user modeling and personalization are non-trivial problems that we investigate.

In this paper, we study user modeling on Twitter and evaluate the quality of user models in the context of recommending news articles. We develop a framework that enriches the semantics of individual tweets and provides a variety of user modeling strategies for constructing semantically meaningful user profiles. The characteristics of these user profiles are influenced by different design alternatives. To better understand how those factors impact the characteristics and quality of the resulting user profiles, we conduct an in-depth analysis on a large Twitter dataset of more than 2 million tweets and answer research questions such as the following:

1. How does the semantic enrichment impact the characteristics and quality of Twitter-based profiles?
2. How do (different types of) profiles evolve over time? Are there any characteristic temporal patterns?
3. How do the different user modeling strategies impact personalization?

2 Twitter-based User Modeling Framework

The proposed user modeling strategies aim to generate user profiles that reflect the interests of a user. Hence, the user profiles will describe to what extent a user is interested in a certain topic. The generic model that can thus be applied for representing user interests can be specified as follows.

Definition 1 *The profile of a user $u \in U$ at a given timestamp $time$ is a set of weighted topics where with respect to the given user u for a topic $c \in C$ its weight $w(u, c, time)$ is computed by a certain function w.*

$$P(u, time) = \{(c, w(u, c, time))|c \in C, u \in U\} \qquad (1)$$

Here, U denotes the set of users while C denotes the set of concepts which represent the topics of interests. To facilitate the interpretation and processing of such user profiles, we typically normalize user profiles so that the sum of all weights in a profile is equal to 1: $\sum_{c_i \in C} w(u, c_i, time) = 1$. With $\vec{p}(u, time)$ we refer to $P(u, time)$ in its vector space model

*The paper on which this extended abstract is based was the recipient of the best paper award of the 2011 International Conference on User Modeling, Adaptation and Personalization [Abel et al., 2011a].

design dimension	design alternatives
topic modeling	(i) hashtag-based, (ii) category-based (iii) entity-based
enrichment	(i) tweet-only-based enrichment or (ii) exploitation of external Web resources
temporal constraints	(i) specific time period(s), (ii) temporal patterns (*weekend*, *night*, etc.) or (iii) no constraints
weighting scheme	(i) TF, (ii) TFxIDF, or (iii) time-sensitive weighting schemes

Table 1: Design space of user modeling strategies.

representation, where the value of the i-th dimension refers to $w(u, c_i, time)$. When designing a user modeling strategy that generates user profiles according to the above definition, there are a couple challenges and design decisions that have to be tackled (see Table 1).

Topic Modeling
What are the actual topics of interests for individual users? According to Definition 1, one thus has to decide what kind of concepts $c \in C$ are used to model topics of interests. We propose three approaches: (i) hashtag-based topic modeling (C is a set of hashtags), (ii) category-based topic modeling (C is a set of 18 broad categories such as sports, politics) and (iii) entity-based topic modeling (C is a set of referenceable entities such as Wikipedia entities).

Enrichment
A core question that we investigate in this paper is whether Twitter messages are a sufficient basis for building user interest profiles that can be applied to provide personalization or whether further enrichment is beneficial to the Twitter-based user modeling. In previous work, we investigate strategies that exploit URLs which the users explicitly posted in their tweets as well as strategies that also aim to link tweets which do not contain a URL with related Web resources [Abel *et al.*, 2011b]. In this paper, we apply these enrichment strategies and analyze whether topics that are extracted from the Web resources such as online news articles, which are linked from the Twitter messages, add value to the user modeling.

Temporal constraints
A third dimension that we investigate in the context of Twitter-based user modeling is given by *temporal constraints* that are considered when constructing the profiles (see Table 1). For example, is it useful to exploit the entire history of a user's Twitter timeline when constructing her user interest profile? Therefore, we first study the nature of user profiles created within specific time periods: we compare profiles constructed by exploiting the complete user history (long-term) with profiles that are based only on Twitter messages published within a certain week (short-term). Second, we examine certain time frames for creating the profiles. For example, we explore the differences between user profiles created on the weekends with those created during the week to detect temporal patterns that might help to improve personalization within certain time frames.

Weighting scheme
Given a topic modeling strategy, an enrichment strategy and a strategy for incorporating temporal constraints, one has to select an appropriate weighting scheme $w(u, c, time)$ that assigns a weight to the concepts/topics of interest (see Def. 1). The weight specifies to what extent a user is interested into the topic c. Our framework provides different weighting schemes ranging from term frequency based methods that count the number of occurrences of c in u's tweets (see TF, Table 1) to more advanced time-sensitive weighting schemes that also incorporate a temporal decay [Abel *et al.*, 2011a; Gao *et al.*, 2011].

Together, the different user modeling building blocks form a rich Twitter-based user modeling framework. By selecting and combining the different design dimensions and design alternatives, we obtain a variety of different user modeling strategies that will be analyzed and evaluated in this paper.

3 Analysis of Twitter-based User Profiles

To understand how the different user modeling design choices influence the characteristics of the generated user profiles, we applied our framework to conduct an in-depth analysis on a large Twitter dataset.

3.1 Data Collection and Data Set Characteristics

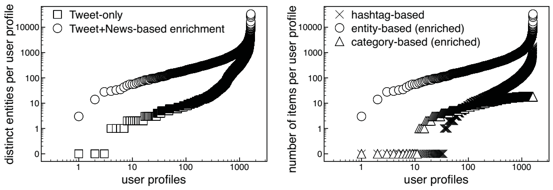
Over a period of more than two months we crawled more than 10 million tweets published by more than 20,000 users. To allow for linkage of tweets with news articles we also monitored more than 60 RSS feeds of prominent news media such as BBC, CNN or New York Times and aggregated the content of 77,544 news articles. As we were interested in analyzing temporal characteristics of the user profiles, we created a sample of 1619 users, who contributed at least 20 tweets in total and at least one tweet in each month of our observation period. This sample dataset contained 2,316,204 tweets.

We processed each Twitter message and each news article to identify categories and entities mentioned in the the tweets and articles. Further, we applied two different linking strategies and connected 458,566 Twitter messages with news articles. For the detailed description and evaluation of the linking strategies we refer the reader to [Abel *et al.*, 2011b]. Our hypothesis is that – regardless whether this enrichment method might introduce a certain degree of noise – it impacts the quality of user modeling and personalization positively.

3.2 Structural Analysis of Twitter-based Profiles

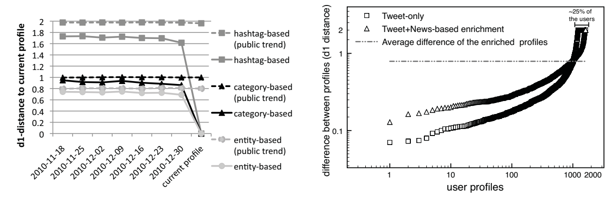
To validate our hypothesis and explore how the exploitation of linked external sources influences the characteristics of the profiles generated by the different user modeling strategies, we analyzed the corresponding profiles of the 1619 users from our sample. In Figure 1, we plot the number of distinct (types of) concepts in the hashtag-, category- and entity-based profiles and show how this number is influenced by the additional enrichment with linked news articles on the Web.

For both types of profiles the enrichment with entities and categories obtained from linked news articles results in a higher number of distinct concepts per profile (see Fig. 1(a)).



(a) Entity-based profiles (b) Comparison of different types of profiles

Figure 1: Comparison between different user modeling strategies with tweet-only-based or news-based enrichment.



(a) Temporal evolution of user profiles (b) Weekday vs. weekend profiles

Figure 2: Temporal analysis of Twitter-based profiles.

Category-based profiles abstract much stronger from the concrete Twitter activities than entity-based profiles. In our analysis, we utilized the OpenCalais [Reuters, 2008] taxonomy consisting of 18 category such as politics, entertainment or culture. The tweet-only-based user modeling strategy, which exploits merely the semantics attached to tweets, fails to create profiles for nearly 100 users (6.2%, category-based) as for these users none of the tweets can be categorized into a category. By enriching the tweets with categories inferred from the linked news articles, we better understand the semantics of Twitter messages and succeed in creating more valuable category-based profiles for 99.4% of the users.

A comparison of the entity- and category-based user modeling strategies with the hashtag-based strategy (see Fig. 1(b)) shows that the variety of entity-based profiles is much higher than the one of hashtag-based profiles. While the entity-based strategy succeeds to create profiles for all users in our dataset, the hashtag-based approach fails for approximately 90 users (5.5%) as the corresponding people neither made use of hashtags nor re-tweeted messages that contain hashtags. Entity-based as well as category-based profiles moreover make the semantics more explicit than hashtag-based profiles. Each entity and category has a URI which defines the meaning of the entity and category respectively.

The advantages of well-defined semantics as exposed by the category- and entity-based profiles also depend on the application context, in which these profiles are used. The results of the quantitative analysis depicted in Fig. 1 show that entity- and category-based strategies allow for higher coverage regarding the number of users, for whom profiles can be generated, than the hashtag-based strategy. Further, semantic enrichment by exploiting news articles which are (implicitly) linked with tweets increases the number of entities and categories available in the profiles significantly and improves the variety of the profiles.

3.3 Temporal Analysis of Twitter-based Profiles

As part of the temporal analysis, we investigate (1) how the different types of user profiles evolve over time and (2) which temporal patterns occur in the profiles. Regarding temporal patterns we, for example, examine whether profiles generated on the weekends differ from those generated during the week. Similar to the click-behavior analysis by Liu et al. [Liu *et al.*, 2010], we apply the so-called d_1-*distance* for measuring the difference between profiles in vector representation: $d_1(\vec{p_x}(u), \vec{p_y}(u)) = \sum_i |p_{x,i} - p_{y,i}|$.

The higher $d_1(\vec{p_x}(u), \vec{p_y}(u)) \in [0..2]$ the higher the difference of the two profiles $\vec{p_x}(u)$ and $\vec{p_y}(u)$ and if two profiles are the same then $d_1(\vec{p_x}(u), \vec{p_y}(u)) = 0$. Figure 2 depicts the evolution of profiles over time. It shows the average d_1-distance of the current user profiles with the profiles of the same users created based on Twitter activities performed in a certain week in the past. As suggested in [Liu *et al.*, 2010], we also plotted the distance of the current user-specific profile with the *public trend* (see Fig. 2(a)), i.e. the average profile of the corresponding weeks.

For the three different profile types, we observe that the d_1-distance slightly decreases over time. For example, the difference of current profiles (first week of January 2011) with the corresponding profiles generated at the beginning of our observation period (in the week around 18th November 2010) is the highest while the distance of current profiles with profiles computed one week before (30th December 2010) is the lowest. It is interesting to see that the distance of the current profiles with the public trend (i) is present for all types of profiles and (ii) is rather constant over time. This suggests (i) a certain degree of individualism in Twitter and (ii) reveals that the people in our sample follow different trends rather than being influenced by the same trends.

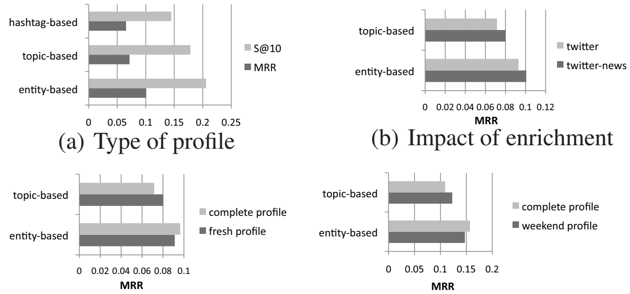
We further investigate how profiles created on the weekends differ from profiles (of the same user) created during the week (see Fig. 2(b)). For category-based profiles generated solely based on Twitter messages, it seems that for some users the weekend and weekday profiles differ just slightly while for 24.9% of the users the d_1-distance of the weekend and weekday profile is maximal (2 is the maximum possible value, see Fig. 2(b)). The news-based enrichment reveals however that the difference between weekend profiles and weekday profiles is a rather common phenomenon: the curve draws nearer to the average difference (see dotted line); there are less extrema, i.e. users for whom the d_1-difference is either very low or very high. Hence, it rather seems that the tweets alone are not sufficient to get a clear understanding of the users concerns and interests.

4 Personalized News Recommendations

In this section, we investigate the impact of the different user modeling strategies on recommending news articles.

4.1 News Recommender System and Methodology

Our main goal is to analyze and compare the applicability of the different user modeling strategies in the context of news



(a) Type of profile (b) Impact of enrichment

(c) Fresh vs. complete profile (d) Weekend recommendations

Figure 3: Results of news recommendation experiment.

recommendations. Therefore, we 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 2 (Recommendation Algorithm) *Given a user profile vector $\vec{p}(u)$ and a set of candidate news items $N = \{\vec{p}(n_1), ..., \vec{p}(n_n)\}$, which are represented via profiles using the same vector representation, the recommendation algorithm ranks the candidate items according to their cosine similarity to $\vec{p}(u)$.*

Given the dataset described in Section 3.1, we considered the last week of our observation period as the time frame for computing recommendations. The ground truth of news articles, which we consider as *relevant* for a specific user u, is obtained via the Twitter messages (including re-tweets) posted by u in this week that explicitly link to a news article. We thereby identified, on average, 5.5 relevant news articles for each of the 1619 users from our sample. We then applied the different user modeling strategies together with the above algorithm and set of candidate items, which contained 5529 items published within the recommendation time frame, to compute news recommendations for each user. The user modeling strategies were only allowed to exploit tweets published before the recommendation period. The quality of the recommendations was measured by means of *MRR* (Mean Reciprocal Rank), which indicates at which rank the first item relevant to the user occurs on average, and *S@10* (Success at rank 10), which stands for the mean probability that a relevant item occurs within the top 10 recommended items of the ranking. We tested statistical significance of our results with a two-tailed t-Test where the significance level was set to $\alpha = 0.01$ unless otherwise noted.

4.2 Results

The results of the news recommendation experiment are summarized in Fig. 3 and validate findings of our analysis presented in Section 3. Entity-based user modeling (with news-based enrichment), which produces according to the quantitative analysis (see Fig. 1) the most valuable profiles, allowed for the best recommendation quality and performed significantly better than hashtag-based user modeling (see Fig. 3(a)). Category-based user modeling also performed better than the hashtag-based strategy – regarding S@10 the performance difference is significant. Since the category-based strategy models user interests within a space of 18 different categories (e.g., politics or sports), it further required much less run-time and memory for computing user profiles and recommendations than the hashtag- and entity-based strategies, for which we limited dimensions to the 10,0000 most prominent hashtags and entities respectively.

Further enrichment of category- and entity-based profiles with categories and entities extracted from linked news articles, which results in profiles that feature more facets and information about users' concerns (cf. Sec. 3.2), also results in a higher recommendation quality (see Fig. 3(b)). Exploiting both tweets and linked news articles for creating user profiles improves MRR significantly ($\alpha = 0.05$). In Fig. 3(c), we further compared strategies that exploited just recent Twitter activities (two weeks before the recommendation period) with the strategies that exploit the entire user history (cf. Sec. 3.3). For the category-based strategy, we see that *fresh* user profiles are more applicable for recommending news articles than profiles that were built based on the entire user history. However, entity-based user modeling enables better recommendation quality when the complete user history is applied. Results of additional experiments [Abel *et al.*, 2011c] suggest that this is due to the number of distinct entities that occur in entity-based profiles (cf. Fig. 1): long-term profiles seem to refine preferences regarding entities (e.g. persons or events) better than short-term profiles.

To examine the impact of temporal patterns, e.g. *weekend pattern* (cf. Sec. 3.3) on the accuracy of the recommendations, we focused on recommending news articles during the weekend and compared the performance of user profiles created just by exploiting weekend activities with profiles created based on the complete set of Twitter activities (see Fig. 3(d)). Similarly to Fig. 3(c) we see again that the entity-based strategy performs better when exploiting the entire user history while the category-based strategy benefits from considering the weekend pattern. For the category-based strategy recommendation quality with respect to MRR improves significantly when profiles from the weekend are applied to make recommendations during the weekend.

5 Conclusions

In this paper, we developed a user modeling framework for Twitter and investigated how the different design alternatives influence the characteristics of the generated user profiles. Given a large dataset we showed that the semantic enrichment enhanced the variety and quality of generated user profiles. Our analysis also revealed that the exploitation of tweet-news relation allows for constructing richer user profiles and has significant impact on news recommender systems. We further analyzed the temporal dynamic of user profiles and discovered that the consideration of temporal patterns such as characteristic differences between weekday and weekend profiles improved the recommendation quality.

Acknowledgements. This work is partially sponsored by the EU FP7 project ImREAL (http://imreal-project.eu).

References

[Abel *et al.*, 2011a] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing temporal dynamics in Twitter profiles for personalized recommendations in the social web. In *WebSci '11: Proceedings of 3rd International Conference on Web Science, Koblenz, Germany*, Koblenz, Germany, 2011.

[Abel *et al.*, 2011b] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Semantic enrichment of Twitter posts for user profile construction on the social web. In *ESWC '11: Proceeding of 8th Extended Semantic Web Conference*, Heraklion, Greece, 2011.

[Abel *et al.*, 2011c] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Supporting website: code, datasets and additional findings. 2011. http://wis.ewi.tudelft.nl/umap2011/.

[Cha *et al.*, 2010] Meeyoung Cha, Hamed Haddadi, Fabrcio Benevenuto, and P. Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM '10: Proceedings of the Fourth International Conference on Weblogs and Social Media*, Washington, DC, USA, 2010.

[Chen *et al.*, 2010] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, Atlanta, GA, USA, 2010.

[Cheng *et al.*, 2010] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *CIKM' 10: Proceedings of the 19th ACM international conference on Information and knowledge management*, Toronto, Canada, 2010.

[Dong *et al.*, 2010] Anlei Dong, Ruiqiang Zhang, Pranam Kolari, Jing Bai, Fernando Diaz, Yi Chang, Zhaohui Zheng, and Hongyuan Zha. Time is of the essence: improving recency ranking using twitter data. In *WWW '10: Proceedings of the 19th international conference on World wide web*, Raleigh, NC, USA, 2010.

[Gao *et al.*, 2011] Qi Gao, Fabian Abel, Geert-Jan Houben, and Ke Tao. Interweaving trend and user modeling for personalized news recommendations. In *WI '11: Proceeding of the 2011 Internation Conference on Web Intelligence*, Lyon, France, 2011.

[Golbeck and Hansen, 2011] Jen Golbeck and Derek L. Hansen. Computing Political Preference among Twitter Followers. In *CHI '11 Proceedings of the 29th International Conference on Human Factors in Computing Systems*, Vancouver, Canada, 2011.

[Hannon *et al.*, 2010] John Hannon, Mike Bennett, and Barry Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*, Barcelona, Spain, 2010.

[Liu *et al.*, 2010] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. Personalized news recommendation based on click behavior. In *IUI '10: Proceeding of the 14th international conference on Intelligent user interfaces*, Hongkong, China, 2010.

[Reuters, 2008] Thomson Reuters. The OpenCalais web service. 2008. http://www.opencalais.com/.

[Sakaki *et al.*, 2010] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW '10: Proceedings of the 19th international conference on World wide web*, Raleigh, NC, USA, 2010.