EgoTR: Personalized Tweets Recommendation Approach

Slim Benzarti[1] and Rim Faiz[2]

[1] LARODEC, ISG University of Tunis, Le Bardo, Tunisia
Slim_benzarti@yahoo.fr
[2] LARODEC, IHEC University of Carthage, Carthage Presidency, Tunisia
Rim.Faiz@ihec.rnu.tn

Abstract. Twitter and LinkedIn are two popular networks each in its territory. Nowadays, people use both of them in order to update their social (Twitter) and professional (LinkedIn) life. However, an information overload problem, caused by the data provided from these two networks separately, troubled many users. Indeed, the main goal of this work is to provide personalized recommendations that satisfy the user's expectations by exploiting the user generated content on Twitter and LinkedIn. We propose a method of recommending personalized tweet based on user's information from twitter and LinkedIn simultaneously. Our Final method considers two main elements: keywords extracted from Twitter and LinkedIn. Those extracted from Twitter are filtered by criteria such as hashtags, URL expansion and Tweets similarity. In order to evaluate our framework performance, we applied our system on a set of data collected from Twitter and LinkedIn. The experiments show that the proposed categorization of the elements is successfully important and our method outperforms several baseline methods.

Keywords: Twitter, LinkedIn, Tweet Recommendation, Content based, Personalization, Skills, and Interests.

1 Introduction

In recent years, Twitter has rapidly become a popular social information network. It is a microblogging platform where short posts are shared between several of users. Recent statistics show that more than 500M users generate more than 300M tweets every day. In other hand, LinkedIn is also a popular social network specialized in the professional user's life that significantly ahead of its competitors and its membership grows by approximately two new members every second. People use social networks not only to maintain social links with other people, but also for several other purposes, as well as sending messages, chatting and gathering news URLs. However, Information Retrieval Systems (IRS) faces new challenges due to the growth and diversity of available data. This huge quantity of information can be exploited and a lot of relevant data can be inferred to answer user's information needs in both social and professional life. In Twitter, the problem of overlapped content is observed when, some relevant tweets are flooded by other ones that might not interest the user at all and which oblige him to seek for the needed information by doing his own scan.

R. Silhavy et al. (eds.), *Intelligent Systems in Cybernetics and Automation Theory*,
Advances in Intelligent Systems and Computing 348, DOI: 10.1007/978-3-319-18503-3_23

In addition, professional interests must not be forgotten, because it is also an important factor to decide whether a tweet is useful or not. Profiling user's personal interests this way may be inaccurate and may not reflect the right user's future intensions. Besides the user's own tweet and retweet history used by many researchers, there are many other kinds of available important information, such as "favorite statues" which is utilized for the very first time to improve the user modeling. In another direction, the user's professional life and his relations with his colleagues can also greatly influence his behavior. By adding LinkedIn as another source of information, recommendations will perfectly satisfy the user expectations and facilitate the task of updating the news concerning his professional interests. In order to fully utilize such information, we propose an egocentric user-based approach exploiting the user generated content extracted from two separate networks to capture personal interests. This method is a promising technology for recommender systems. The recommendation will be based on the observed user posts so that the unobserved user preferences can be inferred from it. In personalized tweet recommendation, tweets are regarded as a set of words, and the user's preferences are obtained by analyzing his interactions.

Our approach exploit first, Twitter features such as tweets, retweet, favorite content, and LinkedIn skills mentioned explicitly by the user to capture his interests in order to create a combined profile which greatly represents his expectations. Then, we will exploit other features to create a tweet profile, composed by a bag of keywords, by combining several criteria such as hashtags, URL expansion and tweets similarity. The use of such features makes our model fully utilize the information mentioned on these two on line networks and do better personalize recommendations Indeed, the recommendation depends on measuring the similarity between the user profile and the tweet profile using the cosine formula. This paper is structured as follows: In Section 2, we survey the related works on personalized tweet recommendation. Section 3 introduces the proposed model. We evaluate our framework in Section 4. Finally, we discuss our work and summarize conclusions and future works.

2 Related Work

2.1 User Personalization

At first, user profile construction is either done in a static way, by gathering information that hardly changes like name, age and so on, or in a dynamic way, by gathering data that frequently changes. User's data are obtained explicitly by the user himself or implicitly by observing his behavior and interactions during his session (history, clicks, pages visited, etc.). According to [18], the user profile contains information such as: 1) Basic information which refers to the name, age, address, etc. 2) Knowledge of the user which is extracted generally from his web page navigation. 3) Interests which are well-defined through a set of keywords. 4) Feedback which design collected information from user's activity and could be deduced from number of clicks and time allowed in consulting resource, etc. 5) Preferences which are characteristics of user describing preferences in specific links or nodes, they could indicated preference in page style presentation, color, etc. [12] defined tags as the means by which users utilize for many purposes like: contributing and sharing, making an opinion, making attention, etc. In [13], the authors discuss the tags usefulness and they conclude that, on the one hand it's used for guiding other users to have information, and on the other hand to receive information about a user due to the history of tagging. [15] analyze the semantics of hashtags in more detail and reveal that tagging in Twitter rather used to join public discussions than organizing content for future retrieval. [17] have defined metrics to characterize hashtags respecting four dimensions: frequency, specificity, consistency, and stability over time. [20] explored the retrieval of hashtags for recommendation and introduced a method which takes in consideration user interests to find hashtags that are often applied to posts related to this topic. [7] Presents an in-depth comparison of three measures of influence, in-degree, re-tweets, and mentions, to identify and rank influential users. Based on these measures, they also investigate the dynamics of user influence across topics and time. Amalthaea, by [21], is one of many systems which create keyword profiles by extracting keywords from Web pages. In fact, the profile is built using one keyword vector for each user and then it is compared to document's vectors using cosine similarity. WebMate by [16] represent user profiles using one keyword vector per user's interest. Twopics introduced by [19], which is an entity-based profiling approach, aims to discover the topics of interest for Twitter users by examining the entities they mention in their Tweets. Other researchers were interested in exploring in depth URL mentioned by the user. [1] Evaluated and showed that the URL expansion strategy achieve 70-80% accuracy. Given the links between tweets and external sources, entities and topics extracted from articles, can be propagated to the corresponding tweets to further contextualize and enhance the semantics of Twitter activities. In this approaches, the authors used only tweets as input to extract the user's topic profile. We suggest including more information such as Tweets, retweets and favorites, which will be taken into account in the detection of user interest. Generally, the main purpose of constructing user's profile in information systems is to adapt the user generated content to answer his expectations.

To Insure a Relevant Profile Construction, We Must Cope with this Problem. The User Profile Evolution Consists Mainly of Apprehending Interest's Change, and Propagating these Changes in the Profile's Representation

2.2 Recommender Systems

2.2.1 Content-Based Recommendation Systems

[24] developed recommender systems that match user preferences (discovered from users profiles), with features extracted from locations (such as tags and categories), to make recommendations. Further recommendation systems provide a user with possible friends based on user's interactions in several social networks [5] and [10]. Other authors use the user location histories that reveal preferences in the friend's recommendation process. Consequently, users with similar past locality have similar preferences and are more likely to become friends. A study on MySpace [2] reveals that user's social connections are related to their geographical belonging. [3] analyze the rating data from Movie-Lens and finds that people at different places have diverse preferences. For instance, people who live in Minnesota are more interested in crime and war movies, whereas users from Florida are more concerned by fantasy and animation movies. To cope with this observation, [4] and [14] have proposed several

algorithms to increase the relevance of the search results. Content recommendations in Twitter aim at evaluating the importance of information for a given user and directing the user's attention to certain items. [8] focus on recommending URLs posted in Twitter messages and propose to structure the problem of content recommendations into three separate dimensions: discovering the source of content, modeling the interests of the users to rank content and exploiting the social network structure to adjust the ranking according to the general popularity of the items.

2.2.2 Collaborative Filtering Recommendation Systems

In [26] a real-time system for on-line web content using a collaborative filtering method was proposed to perform more varied and personalized recommendations within a geographical area. However [8] didn't investigate user modeling in detail, but represent users and their tweets by means of a bag of words, from which they remove stop-words. [6] assumes that most available music recommender systems are based on collaborative filtering methods; they recommend music to user by considering some other user's ratings for the same music pieces. This technique is quite widely utilized, including music shopping services like Amazon or iTunes. However, this recommendation method suffers from the cold start problem. In the proposed approach, three major elements on Twitter are considered: tweet topic level factors, user social relation factors and explicit features such as authority of the publisher, and quality of the tweet. Collaborative filtering approaches exploit information about users who like similar items a long time ago. As discussed in [8], the only issue is that this kind of method obliges each news post to receive instantaneous feedback from numerous users before being recommended to other users. Some other systems rely on textual description of item that could be recommended, for instance, profiles that describe the user's interests regarding the items in the system. Finally, such systems require a means of measuring the compatibility between users and items in order to know which items to recommend to which users. Their results show that using learning to rank over three types of features helped to incorporate real-time web content while further improving the relevance ranking.

2.3 Profiles Merging

Several methods have been developed to construct the user's interests. Three approaches are commonly used [27]:

- **The direct approach**: is to directly ask users what they like, for example, by listing all categories of interests and asking them to make selections.
- **The semi-direct approach**: is to ask users to assign notes to items (e.g. products they have purchased) they have manipulated.
- **The indirect approach**: is to get the user preferences from his friend's votes, by collecting their past ratings concerning a given topic.

[25] have shown that the accuracy of a user profile of a recommender system can be increased by integrating data from other recommender systems. As users often have accounts on different Web 2.0 platforms, the combination of user profiles from different platforms might increase the quality of a user profile as well. [1] combined form-

based and tag-based Web 2.0 profiles. The tag-based profiles were extracted from Flickr, Delicious and StumbleUpon and could not only successfully overcome the cold-start problem but also improve the quality of comprehensive single-platform-based profiles. A significant observation is that the authors detected a very small overlap of the tags that a user used in different systems during the combination of the profiles. This leads to the assumption that users use different social networks for diverse goals: while users use LinkedIn to connect to their business partners, and those who use Twitter might tweet about both private and business related topics.

All the previous presented representations do not consider the time feature which is an important dimension in the user's context. To address this problem, we will attribute a weight which expresses the importance of the keyword and classified them in order to take into account the user interests evolution.

2.4 The Proposed Personalized Tweet Recommendation

The focus of our work is how to extract user profiles from Social Web content, then how to recommend relevant tweets. The problem here is the unstructured nature of the Social Networks data as well as the fact that users do not only interact on Twitter or mention explicitly their skills on LinkedIn related to their work and professional knowledge, but that a huge quantity of tags is often about their social life and other topics.

2.4.1 User Profile Construction

EgoTR is a system for constructing a rich user profile and recommending interesting tweets according to the user interests. Given generated data, EgoTR aims to infer and output relevant recommendations in order to satisfy user expectations.

The profile construction consists in extracting a set of terms that reflects the preferences of the user. First, in order to have a concise representation of the user from LinkedIn, we extract user's information from his profile to process textual data collections into a shape that facilitates discovering knowledge from user mentioned information. The goal is not only to rank or to select user's information, but also to extract salient interests from the user profile in order to build more meaningful and rich representation of their future expectations.

Then, we have to remove from the original user data all information presenting troubles. We proceed in two steps: We filter stop words by eliminating special characters which are very common. Second, we reduce a term to its morphologic root, in order to distinguish the variations of the word itself. Finally, we compute the Term Frequency (TF) to each keyword, consequently a weight will be associated to each one and it best reflects the user's interests. The interests that are returned will be ranked, not only by their frequency, but also by taking into account the number of user's friends who voted for this term. The more the word has votes, the more it is considered important in representing the interests of the user.

Fig. 1. Skills and expertise

The numbers enclosed by the red circles represent the user's friend's ratings or votes (Fig. 1). For the first item "Information Retrieval", 21 persons confirmed that the user is competent in this domain, 13 voted for the second one "Text Mining" and 8 persons confirmed that he is competent in "Natural Language". In our proposed user profiling method, these keywords will have different weights depending on the number of reviews. It should be noted that all used criteria are normalized between 0 and 1.

$$CI_L(i) = W(U; i) * V(i) \quad (1)$$

Where $CI_L(i)$ is the weight of the selected user's interest (Obtained after calculating the TF), *U* represents the user and *i* is the extracted interest. It should be noted that we attributed more importance to this category of keywords, while they are explicitly mentioned by the user and approved by their friends.

EgoTR coordinates input from various sources to enrich user profiles. In order to enhance personalization, we will exploit user generated content on Twitter by analyzing his tweets, retweets and favorites. In which follows, we will detail the followed steps to enrich the user profile. This process involved the use of certain parameters, namely: keywords, hashtags and URLs. We aim to enrich Twitter posts by adding semantic to facilitate browsing the interests behind the user's posts. Furthermore, we retrieved the tweets that were mentioned explicitly by the user as favorite. Twitter users do not only generate contents by posting tweets, but also by sharing links to external resources that can refer to news articles, blogs, company Web pages, other social networks, etc.

To find the words that best represent the semantic content of a tweet, we will use the TF.IDF function. It should be noted that we will attribute different weights to the chosen criteria, since they express different user interests. After performing several combinations using all the selected criteria, we give the formula 2 below which presents the linear combination of the most relevant criteria that we used from Twitter i.e. tweets, retweets, favorites that give the highest value.

$$CI_{TW}(i) = \alpha CI_T(i) + \beta CI_{RT}(i) + \gamma CI_{FAV}(i) \quad (2)$$

The parameters α, β, γ represent the correlation coefficients and their values will be further detailed in the next section: It should be noted that all used criteria are normalized between 0 and 1. After performing the weighting step, a classification according to weight is performed. The user profile that is delivered by EgoTR describes users' characteristics like names, locations or date of birth as well as interests.

2.4.2 Tweet Profile Construction

Our solution is divided into two major steps:

First, we filter tweets containing URLs and extracting keywords that require identification of tweets with common text patterns. Then, once we obtain the URLs and the outputs of the extraction demarche, we abstract uniquely identifiable profile information on the external service like: username or user id. Later, we link the user's Twitter profile to these external services. As a result, we extract keywords which will be considered as interests from the external sources, and as a gain access to more information about the user. Finally, we forward data embedded in URLs to the Twitter profiles.

URL Expansion: To further enrich the semantic of tweets, we implemented a strategy that allows obtaining accurate information about the content of the URLs mentioned by the user. Twitter users do not only generate contents by posting tweets, but also by sharing links to external resources that can refer to news articles, blogs, company Web pages, other social networks, etc. We go in depth and examine the tags that contain URLs in order to infer information about the topic of a post and we assume that those topics are indicative to user's interests.

Tweets Similarity: For a given tweet, we recover the tweets that are similar, and then we classify the results according to their relevance (by calculating the tweets similarity between the obtained tweets and the initial one) to a specific interval of time. It allows taking into account the profile evolution over the time. The similarity between tweets is calculated using the function provided by Lucene. If we take two tweets, and we want to calculate their similarity we rely on sim(t,t').

We denote the current tweet modeled as a vector t:

$$F(\vec{t}, t'') = \frac{\sum_{\vec{t'} \neq \vec{t}} Sim(\vec{t}, \vec{t'})}{|T_t''| - 1} \tag{3}$$

The Hashtags Use: In this step we integrate the use of the same hashtags in order to enrich the tweet's profile. We propose to find tweets that contain similar hashtags and expand the content by detecting the keywords that may be relevant.

$$\text{F(t)} \begin{cases} 1 \; if \; it \; contains \; an \; URL \\ 0 \quad otherwise \end{cases}$$

2.4.3 Merging User Profiles

In this step, we are going to merge the interests previously extracted from both of Twitter and LinkedIn in a single new profile. According to [28], the interests express the level of user attention for a given item by analyzing past interactions. Twitter is known for the vast amount of information exchanged between users, which give us an accurate idea about the user. LinkedIn is a professional social network, where the user mentions relevant information related to his skills and related to their skills and professional profile. In addition to these interests, we will take the votes of the user's friends into consideration. Seen the credibility of LinkedIn and the interest that gives the user who mentions expressly his skills, we will assign more weight to keywords that are part of this network than those extracted from Twitter. Consequently, we will fix α= 0, 4 and β= 0, 6:

$$W_F\,(i) = \alpha\, CI_{TW}(i) + \beta\, CI_L(i) \tag{4}$$

W_F expresses the final weight obtained after combining CI_{TW} which denotes the interest extracted from Twitter. CI_L represents those retrieved from LinkedIn.

3 Tweets Recommendation Based on User Profile

In our work, content-based recommendation is adopted while we rely on a textual description of the tweets and user's profiles that contains the user’s interests. Thus, we will measure the compatibility between user interests, that are extracted from their Twitter interactions and interests mentioned explicitly on LinkedIn, and tweets in order to know which ones are relevant and able to be recommended. To measure the compatibility, we use the cosine function which measures the user-tweet similarity.

The cosine coefficient measures the cosine of the angle between the user profile interests and the tweet profile keywords. It can be computed by normalizing the dot product of the two profiles with respect to their norms where Fti;u denotes the weight of a term in the user profile and Vti;d is its weight in the tweet profile.

$$Cos(\vec{u},\vec{d}) = \frac{\sum_i (F_{t_i,u}.V_{t_i,d})}{\sqrt{\sum_{i'} (F_{t_i,u})^2} * \sqrt{\sum_i (v_{t_i,d})^2}} \quad (5)$$

Content based Content based recommendations aim at evaluating the importance of information for a given user and directing the user's attention to certain items where tweets are regarded as items, and the preferences of users on the tweets are the correlation between users and items. To recommend specific tweet i who is supposed to be the most attractive to the user, the system must find the relative position of the interesting items within the total order of items for a specific user u. To this end, for each user, we aggregate his rankings in the test set by accumulating the weight of the item in order to produce a single total list. The items are again sorted in descending order of their accumulated frequencies. The main goal is to help users to discover new items of interest, therefore we add an additional restriction that the item to be recommended has to be novel for the user, and we remove from the suggestion list all occurrences of the pair (u; i). Finally, we generate a Top-10 recommendation list by selecting the 10 items with the highest score.

3.1 Experiments and Evaluations

Dataset description: We systematically collected tweets through the Twitter API. We retrieved concretely the activity of 7236 users; these people published 31450 tweets, over a two-month period: 1 February, 2013-31 March, 2013 (inclusive). More specifically, we have extracted 20392 tweets, 6389 retweeted statues and 4669 favorites. More details of our dataset are presented in Table 1. We have also retrieved the name, surname, date of birth, occupation, skills and expertise, the votes of the user's friends relative to the 7236 users from LinkedIn through LinkedIn Profile API.

In the following section, we describe the experiment run in the study to evaluate the performance of our approach. To improve the performance of the linear combination, that express the use of different criteria to construct the user profile based on Twitter interactions, we vary the parameters of the proposed equation that allows finding the values of the correlation coefficients α; β; γ that gives the best result.

The formula is

$$CI_{TW}(i) = \alpha\ CI_{T}\ (i) + \beta\ CI_{RT}\ (i) + \gamma\ CI_{FAV}\ (i) \quad (6)$$

and the best correlation coefficients that we will use are:

$\alpha = 0{:}5$
$\beta = 0{:}2$
$\gamma = 0{:}3$.

We can see that the overall performance scores of our proposed features combination of tweets, retweets and favorites gets 68.53 % precision, 64.25 % recall. The obtained results confirm the hypotheses that the combination of tweets, retweets and favorites expresses significantly the user interest.

We observed that the (tweets, retweets and favorites)-based user modeling strategy, performs better than the (tweets, retweets) based method by 4%. However, there is no significant difference in performance between (tweets, favorites)-based with 64.42% and (tweets, Hashtags)-based user modeling strategy 63.55%. We thus find the first evidence for our hypothesis that confirms that the use of favorite's statues criteria, in addition to tweets and retweets, enhances the user modeling better than any other baseline strategy. The result of this combination (68.53%) exceeds the most elevated value of all combinations deemed successful by nearly 5 %. The enriched obtained profile presents a rate of 71.88 % that beyond the profile based only on Twitter interactions by 3.35 %. The evaluation's results confirm that the merge of information from different sources improves the user personalization.

The performance of our system in the recommendation phase depends on the number of interests in the user profile and the number of keywords in the tweet profile. It is noted that the system either provides useful recommendations for the user and bad ones. We notice that when the user profile is poor in terms of interests, erroneous recommendations dominate. As the system acquires more keywords, recommendations become more accurate by answering the exact needs of the user.

From the obtained results, we can announce that the recommendation evolution is based mainly on the number of existing keywords in both profiles (User profile and tweet profile).

3.2 Baseline Approaches Comparison

We used Mean Average Precision (MAP), a popular rank evaluation method to evaluate the proposed approach EgoTR. For a single user, average precision is defined as the average of the P@n values for his tweets, retweets and favorites:

$$AP = \frac{\sum_{n=1}^{N} P@n * RF(n)}{|T|} \quad (7)$$

where n is the number of tweets, |T| is the total number of retweets and favorites criteria for the targeted user, RF(n) is a binary function that describes whether the user has retweeted or checked as favorite the nth tweet in the result list. Finally, MAP can be obtained by averaging the AP values of all the users. We have compared our proposed model to several others baseline approaches.

Retweeted Times: is an objective estimation of the popularity of a tweet. This ranking strategy ignores personalization and assumes that the user's interests are the same as general publics. Profiling: This ranking strategy calculates the similarity between a tweet and the user's profile and shows the tweets sorted by similarity score. In comparison, the EgoTR model assimilates content based models by describing the information as weighted keywords, and gives 0, 6852 MAP. Also, it takes advantages of retrieving interests from Twitter tags and explicit interests from LinkedIn.

According to the previous results, we conclude that our proposed method made a great improvement to tweets recommendation performance. The result can be explained by the fact that the model includes more parameters to describe the personal interests and it is also explained by the terms contained in the profile of the tweet, which help detecting in detail the user's preferences.

4 Discussion

In this section we will discuss the results of P@n (n=1,3,5,10) and MAP on the test set. Chronological strategy gets 0,2287 MAP because, retweeting a status depends on user personal interests more than on the time of the post. The method of ranking by the number of retweeted times, performs poorly with 0,2865 MAP. This means that there is still a wide gap between personal interests and the focus of public attention, which indicates that personalization is very important on Twitter. The profiling method is a classic content based method and gives much better performance with 0,4538 MAP.

Our personalized recommending tweets approach, takes advantage of content filtering based recommendation by extracting contextual information from several online social networks (Twitter and LinkedIn) and incorporating them in our system, our experiments prove that it is helpful for detecting personal interests. The evaluation of our experiments shows the effectiveness of our system regarding problems previously addresses in the first section. The outcome of the described experiments clearly shows the benefits of our EgoTR. Our goal is to prove the importance of our contribution based on extracting useful information from several user profiles. The experiments proved that we got advantage of the term frequency because it reflects how often the user used a term. Second, we noticed that it is important to combine explicit and implicit information mentioned by the user.

The chosen feature combination (tweet, retweet and favorites) allows us proving the effectiveness of our system regarding problems addresses in this proposition. Another point we analyzed during our experiments was the user's activities. In fact, the results show that the more the user interacts with the social systems (Twitter), the more terms our system collects and the more relevant items the profile contains. Consequently, the system allows constructing a rich profile that helps performing a more accurate and targeted personalized tweet recommendation. Thus, we deduced that the use of heterogeneous social annotations, from several sources, provides accurate information for modeling the user profile. Finally, by comparing our system with three baseline systems (Chronological strategy, Retweet strategy and the Profiling strategy), the results reveal the efficiency of our EgoTR approach. We deduce that, according to evaluation metrics Precision and the Mean Average Precision (MAP), our system performs better results.

5 Conclusion and Future Work

In this paper, we have introduced EgoTR, a framework for modeling, enriching, and recommending useful tweets exploiting data available on social networks. Currently, we investigate the extraction of user generated content from both Twitter and LinkedIn that model the users' interests and evaluate them in the context of recommending relevant tweets. We have conduct experiments in the extraction of user data that model his interests. In this work, we rely specifically on Twitter interactions i.e tweets, retweets and favorite statues, then on LinkedIn by analyzing the user mentioned skills. The future directions of this work will focus on gathering data from more than two sources by exploiting other social networks reflecting the user interests and expertise such as CiteULike and dbpedia. We can also provide an opportunity for the user to interact with our system by asking questions that may reflect its interests and their evolutions.

References

1. 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)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng., IEEE Educational Activities Department 17(6), 734–749 (2005)
3. Bianne Bernard, A.L., Menasri, F., Al-Hajj Mohamad, R., Kermorvant, C., Mokbel, C., Likforman-Sulem, L.: Dynamic and contextual information in hmm modeling for handwritten word recognition. IEEE Trans. Pattern Anal. Mach. Intell., IEEE Educational Activities Department 33(10), 2066–2080 (2011)
4. Arase, Y., Xie, X., Duan, M., Hara, T., Nishio, S.: A game based approach to assign geographical relevance to web images. In: WWW, pp. 811–820. ACM (2009)
5. Backstrom, L., Leskovec, J.: Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM 2011, pp. 635–644. ACM, New York (2011)
6. Bischo, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, pp. 193–202. ACM, New York (2008)
7. Cha, M., Mislove, A., Gummadi, K.P.: A measurement-driven analysis of information propagation in the Flickr social network. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 721–730. ACM, New York (2009)
8. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1185–1194. ACM, New York (2010)
9. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 211–220. ACM, New York (2009)

10. Gupta, M., Li, R., Yin, Z., Han, J.: Survey on social tagging techniques. SIGKDD Explor. Newsl. 12(1), 58–72 (2010)
11. Helic, D., Trattner, C., Strohmaier, M., Andrews, K.: on the navigability of social tagging systems. In: Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM 2010, pp. 161–168. IEEE Computer Society, Washington, DC (2010)
12. Hidetoshi, K., Keiji, Y.: GeoVisualRank: a ranking method of geotagged imagesconsidering visual similarity and geo-location proximity. In: Srinivasan, S., Ramamritham, K., Kumar, A., Ravindra, M.P., Bertino, E., Kumar, R. (eds.), pp. 69–70. ACM (2011)
13. Huang, J., Chen, J., Cai, H., Friedland, R.P., Koubeissi, M.Z., Laidlaw, D.H., Auchus, A.P.: In Diffusion Tensor MRI Tractography reveals altered brainstem fiber connections accompanying agenesis of the corpus callosum (2011)
14. Jennings, N.R., Sycara, K., Wooldridge, K.: A roadmap of agent research and development. Autonomous Agents and Multi-Agent Systems 1(1), 7–38 (1998)
15. Laniado, D., Mika, P.: Making sense of twitter. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 470–485. Springer, Heidelberg (2010)
16. Mezghani, M., Zayani, C.A., Amous, I.: and. Gargouri F A user profile modelling using social annotations: a survey. In: Proceedings of the 21st International Conference Companion on World Wide Web, WWW 2012 Companion, pp. 969–976. ACM, New York (2012)
17. Michelson, M., Macskassy, S.: A Discovering users' topics of interest on twitter: a first look. In: Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, AND 2010, pp. 73–80. ACM, New York (2010)
18. Mokrane, B., Dimitre, K.: Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils. In: CORIA, pp. 201–218 (2005)
19. Moukas, A., Moukas, R., Maes, P.: Amalthaea: An evolving multi-agent information filtering and discovery system for the www (1998)
20. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, SIGMOD 2000, pp. 427–438. ACM, New York (2000)
21. Rashid, A.M., Karypis, G., Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach. SIGKDD Explor. Newsl. 10(2), 90–100 (2012)
22. Sandholm, T., Ung, H.: Real-time, location-aware collaborative filtering of web content. In: Proceedings of the 2011 Workshop on Context-Awareness in Retrieval and Recommendation, CaRR 2011, pp. 14–18. ACM, New York (2011)
23. Schubert, P.: and. Koch M. The power of personalization: Customer collaboration and virtual communities. In: Proc. Americas Conf. on Information Systems, AMCIS 2002, Dallas, TX, pp. 1953–1965 (2002)
24. Zhou, L.T.J., Li, M.: User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, pp. 1397–1405. ACM, New York (2011)