A Personalized Tweet Recommendation Approach Based on Concept Graphs

Danae Pla Karidi
IMIS, Athena RC
KDBSL, NTUA
Athens, Greece
danae@imis.athena-innovation.gr

Yannis Stavrakas
IMIS, Athena RC
Athens, Greece
yannis@imis.athena-innovation.gr

Yannis Vassiliou
NTUA
Athens, Greece
yv@cs.ntua.gr

***Abstract*—Twitter users get the latest tweets of their followees on their timeline. In this work we present a tweet recommendation approach, which takes advantage of the semantic relatedness of concepts that interest users. Our approach could be leveraged to build an efficient and online tweet recommender. We construct a Concept Graph (CG), containing a variety of concepts and use graph theory algorithms not yet applied in social network analysis in order to produce ranked recommendations. The usage of the Concept Graph allows us to avoid problems such as over-recommendation and over-specialization, because our method takes into account the true and objective relations between a user's Topics of Interest (ToIs) and the Concept Graph itself. We test our method by applying it on a dataset and evaluate it by comparing the results to various state-of-the-art approaches.**

Keywords—social recommendation; content based recommender systems; concept graph; tweet recommendation;

I. Introduction

Twitter is one of the biggest and most well-known microblogging sites, allowing its users to send and read short messages. Unlike other social media the possible relationships among Twitter users are two, followee or follower. A user becomes a follower when he adds another as a friend while the other will be a followee. Finally when a user publishes a tweet, then this automatically appears on his home page as well as the home pages (user timeline) of his followers.

Twitter is growing rapidly into one of the most popular social network services. Recent statistics show that more than 550 million users generate more than 300 million tweets every day. This leads to a huge quantity of information that can be exploited and a lot of relevant data that can be inferred to answer users' information needs. Let us think of a user's homepage: it is growing every time a followee tweets. The problem is that not all of these tweets are in fact interesting to a user. Moreover this tweet overload complicates the manual tweet retrieval, because of the time required. As a result, users are often tired of searching tweets of interest in their homepage and therefore miss important tweets. A solution to this problem is the development of efficient recommender systems that help users filter the interesting tweets in order to save valuable time from searching among uninteresting tweets and cross passing some actually interesting tweets.

This work has been partially funded by the project "Research Programs for Excellence 2014-2016 / CitySense-ATHENA R.I.C."

Twitter itself has released the @MagicRecs account that sends personalized recommendations as direct messages to a user, if for example a Tweet is favorited or retweeted by a number of user's followees.

In this work we propose a new content based method for personalized tweet recommendation and we present its implementation and evaluation results. The method is based on conceptual relations between users' topics of interest (ToIs). The intuition behind our method is that a user's ToIs are objectively connected in a way that form a user profile. The recommendations based on these connections can avoid the problems of user info availability and overspecialization and can be used to capture the dynamic change of these interests too. The accuracy of the proposed method outperforms in many cases the previous state-of-the art works.

The main goal of our system is to provide a social media user with a new timeline that contains messages that strongly match ones interests and that are not necessarily posted by ones followees. This way the user will not miss messages that are interesting and at the same time the user filters out the non-interesting ones.

The work is structured as follows. Section 2 presents an overview of the current state of the art in areas that are related to tweet recommendation. Section 3 describes the proposed model, while Section 4 presents the experimental evaluation of our model along with a detailed description of the datasets. We conclude the paper in Section 5 with a final conclusion and future work.

II. Related Work

As Twitter is growing in one of the most popular microblogging services, a wide variety of tweet recommendation methods can be found in literature. These methods can be grouped into three categories: Collaborative Filtering, Content-Based and Tweet Ranking methods.

A. Collaborative Filtering

Collaborative filtering (CF) methods make use of the community data in order to build user profiles [1]. The intuition behind these methods is that users that share the same opinion on some topics (interesting, not interesting) tend to have the same opinion on other topics too (user-based CF). Moreover, topics that produce the same opinion from some users tend to receive similar opinions from other users (item-based CF) [2],[3]. Both neighborhood-based methods [4],[5] and model-based methods [6],[12] are subcategories of CF that are used widely in tweet recommendation. Neighborhood-based methods recommend items based on the similarity of user of the item neighbors and model-based methods perform recommendation using matrix factorization model or the probabilistic latent factor model. CF methods use linking data extracted from Twitter like follow and retweet links, in order to construct a network structure. These methods [8],[9] apply network analysis algorithms to the network structure in order to find interesting messages. However, the network construction requires a large volume of link data to be retrieved, stored and analyzed and thus can't be updated in an effective and scalable way when new tweets are published in the stream. Several algorithms and features to be extracted by a user's network have been suggested to identify interesting tweets [10]. Such a feature is the topology of the followers' network [11] that was used in order to recommend users. Collaborative filtering approaches require each tweet to get instantly feedback from numerous users before being recommended to other users, known as the "cold-start" problem. In [12] authors use a model-based method, which proposes online update rules on a stochastic gradient descent style based on the last example observed. In [13] authors propose RMFO-RSV method that maintains a reservoir with a representative set of previously seen data points from the stream, which provides a significant boost in performance compared to the one obtained when only the last example is considered. In [14] the authors use Co-Factorization Machines (CoFM) to address the problem of simultaneously predicting user decisions and modeling content in social media by analyzing rich information gathered from Twitter. These methods consider the relationship between tweets and users and the relationship between users and publishers separately. The problem is that two tweets with the exact same text posted by two users will be evaluated differently, although they have the same content thus they are of the same interest to the user! In [15] the authors present an extending topology based algorithm for recommending users in Twitter. The proposed algorithm classifies the users according to their friendship relations and constructs a class including user ids to recommend the target user. User actions and user mentions are also used to optimize the results. In [16] a collaborative ranking model is proposed, CTR, which considers three major elements on Twitter: tweet topic level factors, user social relation factors and explicit features such as authority of the publisher and quality of the tweet. In [17] the authors propose a probabilistic model based on Probabilistic Latent Semantic Analysis (PLSA) to recommend potential followers to users in Twitter. In [36] the authors propose a methodology to infer interests using some users' followees (topical experts) and social annotations (collected via the Twitter Lists feature).

B. Content Based

A common solution to the cold start and complexity problems is to use other information like the textual content of the items to be recommended [18],[19]. In [20] the authors used crowdsourcing to categorize a set of tweets as interesting or uninteresting and reported that the presence of a URL link is a single, highly effective feature for selecting interesting tweets with more than 80% accuracy. However, this rule may categorize incorrectly an uninteresting tweet (links to meaningless content) as interesting. Content-based methods build user profiles by using the users' history tweets. Such recommenders are often used in domains where a large amount of textual content is available for each user, such as websites. Recommending interesting tweets using content is not easy, because tweets are limited in size. Previous works in content-based methods mainly recommend tweets to users by using content analysis like Latent Dirichlet Allocation (LDA) or TF IDF metrics to represent user's interests. In [21] the authors first created bag-of-word profiles for individuals from their activities and then chose websites most relevant to the profile of the individual as recommendations. [24] and [25] conducted topic modeling of temporally-sequenced documents in Twitter and tried to model the topics continuously over time. These approaches learn topic shifts based on word distributions of tweets, while TS-LDA in [26] the model is learning changes based on topic distributions. Another approach, the Labeled-LDA [27], is used to model a tweet using its labeled information, and then built the probability distribution vector of latent topics to represent the tweet's content. Based on similarity between the topic vectors, the incoming tweets are marked as interesting or not interesting. In [22] authors used Explicit Semantic Analysis [23] to construct the user's interest profile based on Wikipedia concepts, in order to re-rank his timeline. However, in Twitter, the content of users' tweets is much limited and sparse, so that these explicit terms extracted from history tweets are insufficient to reflect user's interests. For example, some latent interests or preferences cannot be characterized in content-based methods [6]. Another approach to analyzing Twitter that uses topics is TwitterRank, which aims to identify influential micro-bloggers [28]. This approach leverages LDA by creating a single document from all user's Tweets and then discovering the topics by running LDA over this "document." Again, such an approach has the problems of LDA since the Twitter data is sparse, and the generated topics are based on terms rather than concepts.

Most of the time, a user's twitter activity is insufficient for creating a reliable profile. For this reason a wide variety of approaches make use of both Content-Based and Collaborative Filtering methods. In [18] authors proposed to create user profiles not from an individual's contents of tweets, but from a group of related individuals' tweets. In [29] authors evaluated a range of different profiling and recommendation strategies, based on a large dataset of Twitter users and their tweets, as well as the relationships between them in order to make useful followee recommendations. In [30] TRUPI is proposed, a system which combines the user social features and interactions and the history of her tweets and also captures the dynamic level of users' interests in different topics to accommodate the change of interests over time. In [31] authors propose a method to predict the probability of a tweet being retweeted based on content features alone. In [39] authors extract features from heterogeneous Open Data in order to recommend the placement of bike stations. Finally, in [37] authors propose Ontology Based recommendations for news recommendations, using a traditional term-based recommender and several semantic-based recommendation algorithms to compare unread news items with the user profile and recommending items with the highest similarity with the user.

C. Tweet Ranking

Some recent approaches focus on recommending tweets from the user's timeline. In [32] the authors use a learning-to rank algorithm that uses content relevance, account authority, and tweet-specific features to rank the tweets in the timeline. Other approaches construct a tweet ranking model making use of the user's retweet behaviour. For example they rank both the tweets and the users based on their likelihood of getting a tweet retweeted [38].

The amount of information provided by Twitter is so large, that most of the already mentioned algorithms become intractable. Many optimization methods were developed to reduce time complexity. For example, in [33] the authors applied clustering algorithms to partition user population, built neighbourhoods for users from the partition, and considered only those neighbourhoods when computing recommendations.

In contrast to collaborative filtering techniques, our approach doesn't face the problem of info availability, due to the fact that it makes no use of twitter linking data. The overall time needed for the method to construct a new user timeline is minimum and thus it can be implemented as a streaming online service too. Our method also avoids the problems of accuracy that are caused due to the limited tweet size, of the LDA or TF-IDF based methods. Our method is using each tweet separately (assigns a topic to each tweet) in contrast to most of content based methods that merge tweets and therefore may miss some topical information. Finally our method takes advantage of the Concept Graph in order to recommend tweets of related topics too (using a specific DFS ranking) and not just the ones of the same topics found in user's tweets. Although ontologies are relevant to the Concept Graph, we believe that concept graphs are less complicated and thus a stable but also lightweight basis for tweet recommendations. Ontologies, when used in social media and more specifically in tweet recommendations (limited content and relations), seem like overkill.

III. Model Description

In this work, we propose a method to construct a new ranked personalized user timeline based on users' interests. In order to accomplish that we construct a user profile exploiting the user's previous tweets. These tweets give us a set of interesting concepts which, when represented in the Concept Graph, form the user's profile. The Concept Graph is represented as concepts extracted from the AlchemyAPI Taxonomy (http://www.alchemyapi.com/), which is a sufficient and compact concept hierarchy. We make use of the Steiner Tree to form user's profile and then use the DFS Post Order tree traversal to rank user's interests. We recommend tweets of content similar to user's profile based on the ranked profile. Despite the fact that we use AlchemyAPI's tools to implement our method, various other tools and taxonomies can be used in its place, because our method's basic principles (Concept Graph, Steiner Tree) are not restricted by the specific tools and taxonomies. Our experiments show that our model is effective and efficient to recommend interesting tweets to users. The experimental results demonstrate its superiority to most of the content based state-of-the art approaches.

A. Intuition

Twitter users read and write content over multiple topics of interest. For example one can tweet about sports, but is also interested in politics and gadgets. As a matter of fact one tweet is relevant to some topic. Moreover, a tweet is usually relevant to a specific topic (a basketball game), and not a topic category (sports). So the user's tweets will reveal a specific aspect of one of his topics of interest. The main intuition behind our method is that a content recommender must take into account the true and objective conceptional connections between a user's specific topics of interest, in order to avoid overspecialization. On the other hand the recommendation of general concepts will lead to over-recommendation of tweets. For example let's assume a user is interested in ancient history. A recommender of general topics would recommend tweets about "science" that include chemistry, computer science, medicine, etc. Our approach tries to deal with this problem by ranking the inferred user's interests in a way that both over-recommendation and overspecialization can be avoided.

Our approach consists of the following basic steps:

- Construction of the user's interest profile (based on the Concept Graph) using the semantic information retrieved from her tweets as described in section User Profile
- Assignment of a topic (contained in the Concept Graph) to every tweet as described in section Tweet Representation
- Usage of graph theory algorithms (on the Concept Graph) in order to calculate tweet relatedness to users' profiles and to rank them in descending order of interest

B. Concept Graph

Our method is based on the Concept Graph (CG), which is a compact graph whose nodes represent concepts and edges represent relations between them. In our work the Concept Graph is a way to exploit logical relations between topics of interest in order to provide interesting and efficient tweet recommendations.

In order to construct the CG we use AlchemyAPI service, which offers twelve API functions as part of its text analysis services that use natural language processing techniques to analyse text content. We use the AlchemyAPI Taxonomy service and its Categories dataset, which is a set of concept categories and subcategories extended up to 5 levels deep. For example the category of "music genres", which is a subcategory of "music", has 17 subcategories each of them represents a music genre. This example is shown in Fig. 1.

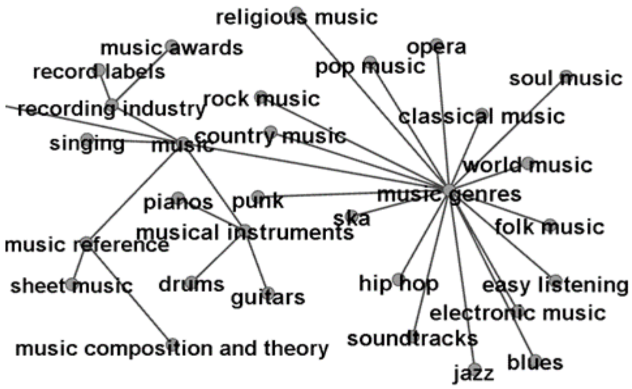

Figure 1

We can regard all Alchemy Taxonomy concepts as a link undirected graph G= (V, E), where V= {vi} is the concept set from AlchemyAPI Taxonomy. Each concept vi is connected with the concept vj, if and only if these concepts are related in AlchemyAPI Taxonomy Dataset via an edge that belongs to the set E= {ei} of graph edges. All edges are of equal weight.

The CG constructed in this way is undirected and consists of 1092 nodes (concepts) and 1323 edges (concept relations). As concept relations we use the relation category-subcategory, and the existence of an edge between two concepts is an indicator of relatedness. The CG covers the vast majority of concepts that are used in everyday life, therefore it provides a wide knowledge base for our recommender (users profile construction).

The main reason of creating the Concept Graph is that we want a common basis to compare user's topics of interest. Concept Graph is thus a scientific objective basis that semantically outperforms the LDA self-topic approaches, as well as term frequency ones that lack in efficiency due to tweets' small size in terms of word count.

We assume that our method uses a representation of the user's interest profile which will be more accurate and solid, avoiding conceptual overlaps. At the same time our method gives the opportunity to develop a lightweight implementation, thus the opportunity to build an efficient online application.

C. *Tweet Representation*

In order to assign topics to tweets we use AlchemyAPI's Taxonomy API, which is an online service for semantic text analysis using natural language processing. The service automatically categorizes text and HTML into a hierarchical concept taxonomy. Using complex statistics and natural language processing technology, the Taxonomy API is able to classify a tweet into its most likely topic category. The category set is the same that was used to build the Concept Graph.

D. *User Profile*

In this section, we present a new model to build a user profile by analysing the tweets of the user, making use of the Concept Graph. The user profile is made based on a set of recent tweets. Each of these history tweets are assigned to a main topic concept using the method presented in the previous section. In order to get the related concepts, we find the Steiner Tree from the Concept Graph containing the extracted topics, which are mentioned as topics of interest. The intuition behind using Steiner Tree is as follows. If another concept is connecting two or more topics of interest in the Concept Graph, then this concept is likely itself a topic of interest. Some topics may not be directly related to one's interests. However, if a concept belongs to the Steiner Tree of the topics of interest, then the likelihood of it being itself a topic of interest increases. This is so because the Steiner Tree provides an optimum set of nodes and edges to connect the topics of interest.

E. *Steiner Tree*

The minimum Steiner Tree [35] describes a way to connect a set of nodes in the "cheapest" way. Steiner tree problem is similar to the minimum spanning tree problem: given a set V of points (vertices), connect them by a new graph of shortest length, where the length is the sum of the lengths of all edges. In our work the edges are all of the same length. The difference between the Steiner tree problem and the minimum spanning tree problem is that, in the Steiner tree problem, extra intermediate vertices and edges may be added to the graph in order to reduce the length of the spanning tree. These new vertices (concepts) introduced to decrease the total length of connection are known as Steiner vertices.

F. *Ranking DFS Postorder*

Our goal to provide a personalized user timeline, requires that the recommended tweets are presented ranked in order of interestingness. The Concept Graph contains concepts in 5 level depth relativity. Recommending high level topics, we are expecting to get too many false positives (too many recommendations that are false – too abstract). We assume that a user is interested in reading tweets about specific topics and in order to avoid over-recommendation we use a DFS traversal of the user's Steiner Tree.

Our implementation of the proposed ranking method, which we named TGS-post, results in the Personal Interest Tree and mainly consists of three steps:

- Retrieve Twitter Stream every n minutes
- Filter the tweets whose topic is included in user's Steiner Tree
- Rank those tweets based on the DFS traversal of the Steiner Tree beginning from the most frequently occurring node in user's tweets

G. *Alternative Ranking Methods*

In order to optimize our system's efficiency, we designed and implemented two alternative ranking methods:

- TGS-Freq: After filtering the tweets whose topic is included in user's Steiner Tree, in order to get the user's Personal Interest Tree, we rank them based on the frequency of their occurrences in user's tweets

- TGS-BFS: After filtering the tweets whose topic is included in user's Steiner Tree, in order to get the user's Personal Interest Tree, we rank them based on the BFS traversal of the Steiner Tree

H. Example

In order to understand the whole process, we give an example. After we construct the CG (Section III B), we pick the user Paul Mason widely known as an economic and politics journalist and broadcaster, currently working as economics editor at Channel 4 News. We retrieve his timeline, and get the topics of interest, using AlchemyAPI (Section III C), which are shown in Table I.

Table I

Post	Type	Topics Of Interest
"Reglingnoics"!Https://t.co/q5lozpnd4l In other respects "had so many strange ideas"! Greekdept	Retweet	Annual report, statistics
According to Regling I had "strange ideas" http://t.co/Qwp4ww8d9 H Read them here & compare them to Mr Regling's: https://t.co/Qmz0uNYuT3	Tweet	Politics, finance, society
Tonight on BBC One's Question Time http://t.co/wghmmx4grh	Tweet	Radio
Il Fatto Quotidiano interview on M. Renzi's 'comment' & my reply http://t.co/fhfmidntzd	Tweet	Reading
Interviewed by POLITICO on the 3rd Bailout, Schauble-Merkel, the Eurozone & the refugee crisis http://t.co/hpuyrmvggw	Tweet	Lobbying
DER SPIEGEL: Complicit in Corruption: How German Companies Bribed Their Way to Greek Deals http://t.co/9cdikw4ahn	Tweet	Government
Ne pas manquer vendredi soir le prochain live de Mediapart: deux heures avec https://t.co/xtktgivrba	Retweet	Tech news
The lenders are the real winners in Greece http://t.co/ltjtyvu5zw	Tweet	Government
We would like to see you again fighting together for Syriza victory...Your division it's a defea…	Retweet	Business and industrial
The Double Purpose of These Elections http://t.co/phl26kcmf2	Tweet	Elections
Paul Krugman on Greece's 3rd mou: An agreement designed to fail http://t.co/hhwiw5fgop	Tweet	Government, greece
Yanis Varoufakis - 'Left should beware of friends who fear confronting the rich' http://t.co/posnmlqviq	Retweet	Unions
Greece should follow Germany's WALK not its TALK: my PBS article http://t.co/ticy8kg1vk but Varoufakis' plan ignored https:…	Retweet	greece

Following we get the Steiner Tree of Concept Graph containing the Topics of Interest (Section III D-E), which is shown in Fig. 2. In the sequel, we present the user's Personal Interest Tree after the DFS Post Order ranking, beginning from the most frequently occuring topic ("government") : government, law, govt and politics, politics, lobbying, elections, business and industrial, science, mathematics, statistics, company, annual report, finance, society, work, unions, art and entertainment, radio, hobbies and interests, reading, technology and computing, tech news, travel, tourist destinations, Greece

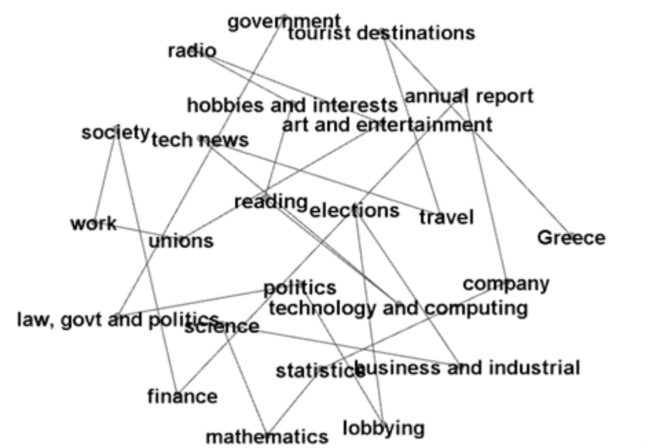

Figure 2

Finally, we filter the streaming tweets that belong to topics included in the Steiner Tree (Fig. 2) and insert them in the new timeline according to their ranking from the DFS Post Order traversal (Section III F).

IV. EVALUATION

In order to evaluate our method we conducted an offline evaluation test and compared it with the most popular state-of-the art methods. For a set of users we gathered their most recent tweets, constructed their profiles, and recommended tweets based on the approach described in Section 3. Our user dataset was constructed crawling tweets and retweets from "The Twitter 100" users of 2012 (The Twitter 100: Britain's titans of the Twittersphere - www.independent.co.uk) list. This is a list of Britain's most influential users of 2012 based on PeerIndex that measures interactions across the web to help users understand their impact in social media. Using Twitter's API, we crawled the twelve most recent tweets for each user in the list. This dataset was enriched with the users' profiles (Personal Interest Trees). Specifically the methodology of the Personal Interest Tree construction was for each user:

- Crawl his twelve most recent tweets (twelve was chosen to avoid scalability and info availability issues)
- Assign a topic (ToI) to every tweet (as described in section Tweet Representation)
- Extract the Steiner Tree from the Concept Graph containing the ToIs
- Execute a DFS traversal of the user's Steiner Tree regarding as root node the ToI that was assigned more frequently to the tweets (most frequently assigned ToI)

The resulting tree is the user's Personal Interest Tree. Finally, our dataset consists of a hundred users and their Personal Interest Trees, which are represented as vectors of ToIs ordered according to the DFS traversal. Subsequently, we constructed a test set in order to evaluate our method. We decided to build this test dataset out of the users' retweets, because we assume that when a user retweets a post, he is most likely interested in it. Tweet replies were not used here to avoid bias, due to long or personal Twitter conversations. Then we assigned a ToI to each retweet (as described in section Tweet Representation). The test dataset consists of the most recent retweets crawled from the users' timelines (500 retweets) and their Topics of Interest. The test process was made in the following stages for each user in the first dataset:

- Get user's Personal Interest Tree
- For each ToI in the vector (beginning from the first) get all retweets of the same ToI from the test dataset
- Store them in the recommendation list
- Continue from stage 2 for the next ToI in the Personal Interest Tree vector

This way we manage to rank all retweets from test dataset according to user's Personal Interest Tree vector and store them in the recommendation list. Subsequently, we computed four performance measures: precision-at-k, accuracy-at-k mean average precision and average accuracy. Precision-at-k corresponds to the precision (information retrieval performance measure) calculated in the first k recommendations in the recommendation list. Accuracy-at-k corresponds to the accuracy (information retrieval performance measure) calculated in the first k recommendations in the recommendation list. In both measures as relevant elements we consider the retweets made by the user (e.g. all user's retweets that exist in recommendation list are considered as true positives). We conducted experiments from k =1 to 10. The results are shown in Tables II, III.

Table II

k	Accuracy-at-k	Precision-at-k
1	0.992357841362	0.236559139785
2	0.989947299479	0.195652173913
3	0.988908567811	0.195652173913
4	0.988507026029	0.16847826087
5	0.988767690328	0.154347826087
6	0.988540217397	0.143115942029
7	0.988148831155	0.143115942029
8	0.987761826312	0.126358695652
9	0.987775232302	0.115942028986
10	0.987828519001	0.105434782609

Table III

Mean Average Precision	**0.157973978161**
Average Accuracy	**0.988854305118**

As we can see in Fig. 3 and Fig. 4, our method reaches a mean average precision score of 15.7% while the average accuracy is 98.8%. This means that our recommender can successfully retrieve the interesting (retweeted) and not interesting tweets (true positive and true negative results), but still recommends some tweets that were not retweeted by the user (false positives). We can also observe that our model outperforms in terms of precision and accuracy the most common state-of-the art methods (lda, tfidf-simple, tfidf-word pairs, muifuot [34]) introduced in section Related Work.

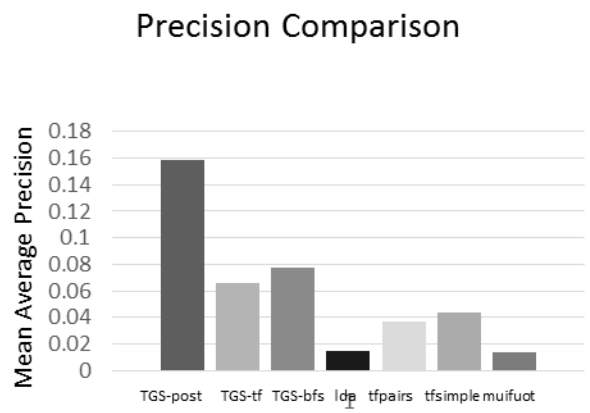

Figure 3

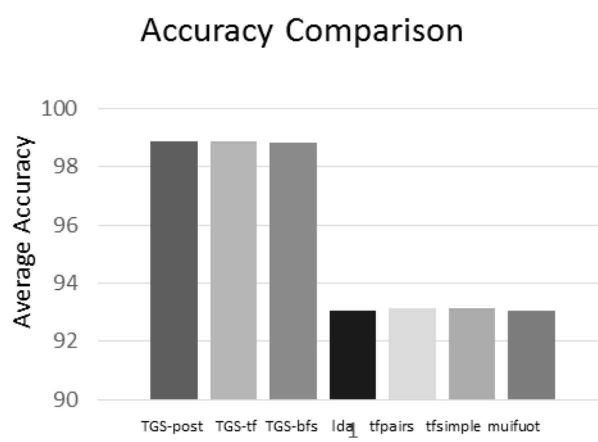

Figure 4

V. Conclusion and Future Work

In this work we propose a new content based method for personalized tweet recommendation, based on conceptual relations between users' topics of interest and we present its implementation and evaluation results. The method takes advantage of the objective relation between user's ToIs (Topics of interest) and a Concept Graph. The recommendations based on these connections can avoid the problems of user info availability, overspecialization and can be used to capture the dynamic change of these interests too in a scalable way. Our experiments with real-life data sets have demonstrated the effectiveness in tweets recommendation.

Some future research directions are the following. Regarding the Concept Graph we may reconstruct it using more types of relations between concepts (not only the relation category-subcategory). Moreover, for the recommendation algorithm, we can use weights on the user's profile (ToIs graph) based on the "amount of interest" that a user has concerning each topic. This will allow us to test different network analysis algorithms (instead of just the Steiner Tree) to extract the recommended ToIs. Finally, we plan to develop a full web application tool based on the proposed method in order to use it as an online twitter service.

REFERENCES

[1] Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in artificial intelligence 2009 (2009): 4.

[2] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." Knowledge and Data Engineering, IEEE Transactions on 17.6 (2005): 734-749.

[3] Schafer, J. Ben, et al. "Collaborative filtering recommender systems." The adaptive web. Springer Berlin Heidelberg, 2007. 291-324.

[4] Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." Proceedings of the 10th international conference on World Wide Web. ACM, 2001.

[5] Shi, Yue, Martha Larson, and Alan Hanjalic. "Exploiting user similarity based on rated-item pools for improved user-based collaborative filtering." Proceedings of the third ACM conference on Recommender systems. ACM, 2009.

[6] Koren, Yehuda. "Factorization meets the neighborhood: a multifaceted collaborative filtering model." Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008.

[7] Rendle, Steffen. "Factorization machines with libFM." ACM Transactions on Intelligent Systems and Technology (TIST) 3.3 (2012): 57.

[8] Romero, Daniel M., et al. "Influence and passivity in social media." Machine learning and knowledge discovery in databases. Springer Berlin Heidelberg, 2011. 18-33.

[9] Yang, Min-Chul, et al. "Finding interesting posts in twitter based on retweet graph analysis." Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012.

[10] Lauw, Hady W., Alexandros Ntoulas, and Krishnaram Kenthapadi. "Estimating the quality of postings in the real-time web." Proc. of SSM conference. 2010.

[11] Armentano, Marcelo G., Daniela Godoy, and Analía Amandi. "Topology-based recommendation of users in micro-blogging communities." Journal of Computer Science and Technology 27.3 (2012): 624-634.

[12] Rendle, Steffen, and Lars Schmidt-Thieme. "Online-updating regularized kernel matrix factorization models for large-scale recommender systems." Proceedings of the 2008 ACM conference on Recommender systems. ACM, 2008.

[13] Diaz-Aviles, Ernesto, et al. "What is happening right now... that interests me?: online topic discovery and recommendation in twitter." Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

[14] Hong, Liangjie, Aziz S. Doumith, and Brian D. Davison. "Co-factorization machines: modeling user interests and predicting individual decisions in twitter." Proceedings of the sixth ACM international conference on Web search and data mining. ACM, 2013.

[15] Yigit, Melike, Bilal E. Bilgin, and Adem Karahoca. "Extended topology based recommendation system for unidirectional social networks." Expert Systems with Applications 42.7 (2015): 3653-3661.

[16] Chen, Kailong, et al. "Collaborative personalized tweet recommendation." Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012.

[17] Kim, Younghoon, and Kyuseok Shim. "Twitobi: A recommendation system for twitter using probabilistic modeling." Data Mining (ICDM), 2011 IEEE 11th International Conference on. IEEE, 2011.

[18] Balabanović, Marko, and Yoav Shoham. "Fab: content-based, collaborative recommendation." Communications of the ACM 40.3 (1997): 66-72.

[19] Schein, Andrew I., et al. "Methods and metrics for cold-start recommendations." Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2002.

[20] Alonso, Omar, et al. "Detecting uninteresting content in text streams." SIGIR Crowdsourcing for Search Evaluation Workshop. 2010.

[21] Pazzani, Michael J., Jack Muramatsu, and Daniel Billsus. "Syskill & Webert: Identifying interesting web sites." AAAI/IAAI, Vol. 1. 1996.

[22] Lu, Chunliang, Wai Lam, and Yingxiao Zhang. "Twitter user modeling and tweets recommendation based on wikipedia concept graph." Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.2012.

[23] Gabrilovich, Evgeniy, and Shaul Markovitch. "Computing semantic relatedness using wikipedia-based explicit semantic analysis." IJCAI. Vol.7.2007.

[24] Kawamae, Noriaki. "Trend analysis model: trend consists of temporal words, topics, and timestamps." Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011.

[25] Wang, Xuerui, and Andrew McCallum. "Topics over time: a non-Markov continuous-time model of topical trends." Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.ACM,2006.

[26] Yang, Min-Chul, and Hae-Chang Rim. "Identifying interesting Twitter contents using topical analysis." Expert Systems with Applications 41.9 (2014):4330-4336.

[27] Ramage, Daniel, Susan T. Dumais, and Daniel J. Liebling. "Characterizing Microblogs with Topic Models." ICWSM 10 (2010): 1-1.

[28] Weng, Jianshu, et al. "Twitterrank: finding topic-sensitive influential twitterers." Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010.

[29] Hannon, John, Mike Bennett, and Barry Smyth. "Recommending twitter users to follow using content and collaborative filtering approaches." Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010.

[30] Elmongui, Hicham G., et al. "TRUPI: Twitter Recommendation Based on Users' Personal Interests." Computational Linguistics and Intelligent Text Processing. Springer International Publishing, 2015. 272-284.

[31] Naveed, Nasir, et al. "Bad news travel fast: A content-based analysis of interestingness on twitter." Proceedings of the 3rd International Web Science Conference. ACM, 2011.

[32] Duan, Yajuan, et al. "An empirical study on learning to rank of tweets." Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010.

[33] Sarwar, Badrul M., et al. "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering." Proceedings of the fifth international conference on computer and information technology. Vol. 1. 2002.

[34] Pennacchiotti, Marco, et al. "Making your interests follow you on twitter." Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

[35] Hwang, Frank K., Dana S. Richards, and Pawel Winter. The Steiner tree problem. Vol. 53. Elsevier, 1992.APA

[36] Bhattacharya, Parantapa, et al. "Inferring user interests in the twitter social network." Proceedings of the 8th ACM Conference on recommender systems. ACM, 2014.

[37] IJntema, Wouter, et al. "Ontology-based news recommendation." Proceedings of the 2010 EDBT/ICDT Workshops. ACM, 2010.

[38] Uysal, Ibrahim, and W. Bruce Croft. "User oriented tweet ranking: a filtering approach to microblogs." Proceedings of the 20th ACM international CIKM. ACM,2011

[39] Chen et al. "Bike sharing station placement leveraging heterogeneous urban open data." Proceedings of the UbiComp2015. ACM,2015.