IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018

www.ijasrm.com

ISSN 2455-6378

A Cohesive Framework for Providing Recommendations In Microblogging Systems

Radhakrishnan Indra[1], Muthuraman Thangaraj[2]

[1]Manonmaniam Sundaranar University,Government Arts College for Women,
Affiliated to Alagappa University,
Sivagangai-630562, Tamilnadu, INDIA.

[2]Madurai Kamaraj University, Madurai, Tamilnadu, INDIA

Abstract

Microblogging is a broadcast medium that allow users to exchange small sentences, individual images or video links. These small messages are referred to as Microblogs or tweets. There are a large number of real-time microblog messages generated every day which results in information overload. Many tweets are marked with Hashtags, which usually represent annotation or metadata for describing microblogs. In microblogging three entities are involved such as Microbloggers : someone who uses a microblog, Microblogs and Hashtags. In this microblogging system if all these three recommendations are provided then the microblogging becomes the personalized microblogging. Existing recommendation systems provide anyone recommendation only. But this paper proposes the recommendation model Recommendations in MicroBlogging Systems providing all the three recommendations in microblogging systems using Semantic Web.

Keywords: *Recommendation, Microblogging System, Microblogger, Microblog, Hashtag, Semantic Web.*

1. Introduction

Recommender systems make use of different sources of information for making predictions and subsequently provide recommendations. These recommendations are useful to balance factors like accuracy, novelty, dispersity and stability. Recommender systems have developed in parallel with the web. Recommender systems are solution to overcome information overloading.

A microblog is a form of a blog where people send short messages of text or media such as pictures, video, or sounds. These messages can be sent to either a small group or to the public through a website. Microblogs can be sent via different sources including cell phone text messages, email and instant messages through a website. A microblog has much smaller pieces and an entry of it can be just one sentence, or a link to an image or short video.

Microblogs are in use in many ways on different websites. Status updates feature provided by many social networking sites (such as Facebook or MySpace) where the person tells their friends what they are thinking or doing is also a kind of microblogging. Microblogs that exist to promote websites, services and products, and to promote collaboration within an organization are commercial microblogs. Users can control who can read their microblogs using features such as privacy settings available in some microblogging services. Some newer sites allow us to share media like videos, pictures or sounds directly instead of sending a link. Many microblogging services are available such as Gab, Twitter, Twister, Tumbir, Yammer, Tout etc.

In microblogging systems users will be following others and the user will be followed by others. If a user has many followees, then the user will struggle with information overload. In this case Microblogger recommendation will be useful by recommending suitable followees to the user, so that the user can select the followee list. Many tweets are marked with Hashtags, which usually represent annotation or metadata for describing microblog posts. These Hashtags can be used for personal organization, easy retrieval and finding related microblogs or groups. As these Hashtags are freely chosen keywords there are problems with ambiguity

such as synonymy, polysemy, singularity, plurality, acronym, abbreviations etc. To overcome these problems Hashtag recommendation will be useful by providing meaningful Hashtags. Due to the information overload and its varieties in microblogging it is difficult to find out the relevant Microposts by the users. Therefore, recommender systems play an important role in filtering and customizing the desired microposts to the users. From the discussion it is clear that in microblogging services, all the three Microblogger, Microblog, Hashtag recommendations plays a vital role. But existing systems provide either Microblog or Microblogger recommendation and some systems provide hashtag recommendation. So there is a need for a system that provides all these three recommendations. Hence, this paper presents the cohesive framework for providing Microblogger, Microblog and Hashtag recommendations in microblogging sites using Semantic Web (SW).

SW is an evolving development of the World Wide Web in which meaning (semantics) of information and services on the web is defined, making it possible for the web to understand and satisfy the requests of people and machines to use the web content.

Rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes the proposed architecture that is capable for providing all three recommendations. Section 4 presents the result of the proposed system and section 5 concludes the paper.

2. Related Work

In this section some of the research literature related to each of the Microblogger, Microblog or Hashtag recommendation is given.

Given a user query a ranked list of microbloggers is produced by a content-based microblogger recommendation model (Burak Celebi H & Uskudarli S, 2012). The user query has been enriched with related keywords. Then using various forms of queries candidate microbloggers are fetched and their contents are processed. Content-based scores of candidate microbloggers are found to recommend and rank them. However, disambiguation of user queries and content is not taken into consideration. This can be resolved through Semantic Web resources and techniques.

To analyze users' possible behaviors and predict their potential friends in microblogging a temporal-topic model is proposed (Zheng, Nan & Song, Shuangyong & Bao, Hongyun, 2015). The model learns users' latent preferences by extracting keywords on aggregated messages over a period of time via a topic model and then the impact of time is considered to deal with interest drifts. Finally users' potential interests on others is predicted based on the sequence of users' interests along the timeline. Social relationships among users (i.e., followers, followees) can be used to improve the performance of this system.

For followee recommendation in microblog, a system is proposed using matrix factorization model with structural regularization (Yan Yu and Robin G. Qiu, 2014) and a recommendation of the micro-blog friend recommendation algorithms, which has two broad categories and three types: the recommendation algorithm based on content, the topology recommendation algorithm based on social relations and the filtering recommendation algorithm is made (Yang L et al., 2018).

An improved Apriori-based algorithm is proposed to provide a recommendation for users with a common interest in Microblog (Lin L et al., 2018).

A system utilizes the sentimental information to help with the microblog recommendation (Wenjuan Cui et al., 2017). A sentiment classifier is built based on the contextual information of the microblogs, the sentimental feature set is acquired and then a latent factor model incorporating the sentimental features and other information in microblogs is designed.

A scholarly microblog recommendation system is proposed to benefit scientific research (Yang Yu, Xiaojun Wan & Xinjie Zhou, 2016). This system automatically collects and mines scholarly information from Chinese microblogs, and makes personalized recommendations to researchers. It uses two different neural network models which learn the vector representations for both users and microblog texts. Then the recommendation is accomplished based on the similarity between a user's vector and a microblog text's vector. Nevertheless, in this method latent associations are not revealed and only one recommendation is provided.

A system Topic-Specific Translation Model [TSTM] is proposed to suggest hashtags for microblogs (Zhuoye Ding, Qi Zhang & XuanJing Huang, 2012). This system combines the advantages of both topic model and translation model. Though this system

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

makes Hashtag recommendation based on the user interaction, it may not be of interest to some users.

To provide time-aware personalized hashtag recommendation on social media a method extending the translation based model and incorporating the temporal and personal factors is proposed (Qi Zhang et al., 2014). Here, only the words are taken into consideration and not the concepts. So there is no semantic grounding.

Hashtag recommendation for multimodel microblog using co-attention network is proposed (Qi Zhang et al., 2017).

Hence, this paper proposes a cohesive framework, Recommendations in MicroBlogging Systems (RMBS) for providing all the three recommendations using semantic web. It matches each microblog post with the Context and predicts new associations among Microblogger, Microblog and Context.

3. Proposed RMBS Architecture

The proposed RMBS architecture is shown in figure 1. This system extracts users' microblog posts from the microblogging site. After preprocessing the microblog posts, the microblog posts are matched with the context and the three entities Microblogger, Microblog and Context in the microblog post are represented in Tensor. Latent semantic associations among three entities Microblogger, Microblog and Context are revealed using Latent Semantic Indexing (LSI) based on Randomized Singular Value Decomposition (RSVD). The resultant core Tensor with weights is stored in the database. Ontology is constructed from the data available on the database. Finally recommendations are performed using Simple Protocol And RDF Query Language (SPARQL). All the processes except recommendation is performed offline.

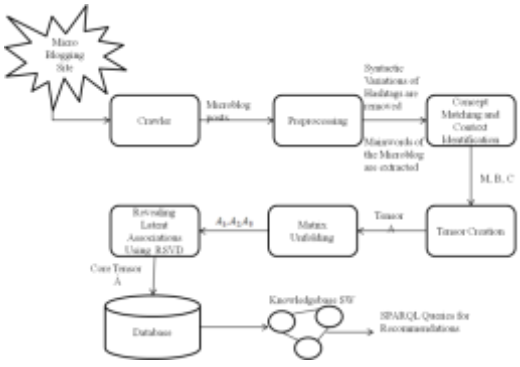

Fig. 1 RMBS Architecture

3.1 Crawler

Microblogs posted or reposted by the users are extracted from the Microblogging site. These microblog posts are in the form of quadruples <Microblogger, Microblog, Hashtag, Time>. Time refers to the time at which the blog was posted by the user. Following notations are used to represent each of the four entities in this quadruple.

M – Microblogger, B – Microblog, H – Hashtag, T – time

These quadruples are stored in the database with one record for each post. The Microblogger and Microblog are represented using unique numeric id. The three entities Microblogger, Microblog and Hashtag in the time interval of a week are taken from the database for further processing. Users' interests change as time goes by. To catch the users' dynamically changing interests, User's Microblog posts are to be analyzed every week. This is the reason for taking only microblog posts of users within a week time.

3.2 Preprocessing

If a microblog post has a Hashtag/Hashtags then it is mapped directly to a concept after finding the base word using stemming. Otherwise the POS (Part of Speech) segmentation of a Microblog is performed. Stop words like articles, pronouns and symbols are removed. Each Microblog is represented as a bag of words having nouns, abbreviations, idioms and academic vocabularies in the blog.

3.3 Concept Matching and Context Identification

Each of the words of the Microblog is mapped directly to the concepts using WordNet. In WordNet each concept is associated with a synset. synset is a synonym set containing synonymic terms which represent the meaning of the concept. The set of concepts for a word 'w' is represented as concept(w) = { c | w ∈ synset(c)}. If concept of a word is not found by direct mapping then partial mapping is performed. If partial mapping is also not able to catch the concept of a word then term mapping is performed (Endang Djuana, Yue Xu & Yuefeng Li, 2011). Now each Microblog is represented by bag of concepts instead of words. Among these concepts the best one that suits the Microblog has to be chosen which will represent the context of the Microblog. For this purpose measure

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

of dominance (Sana Hamdi et al., 2012) (Eq. 1) is used. Dominance is used to compute the score of each concept corresponding to which extent the concept dominates the other ones.

Dominance(concept) =

$$\frac{No.of\ words\ corresponding\ to\ this\ concept}{Total\ no..of\ words\ in\ the\ microblog} \quad (1)$$

The dominant concept , i.e., the one that presents the highest appearance frequency, will be considered as the context of the Microblog. After matching each microblog with the context, the triples <M, B, H> becomes <M, B, C>, where C is the context of the microblog. Fig. 2. shows it pictorically.

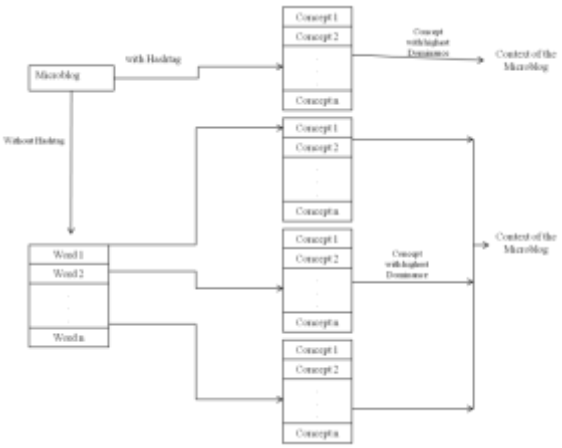

Fig. 2 Concept Matching and Context Identification

3.4 Tensor Creation

A multidimensional matrix is called as a Tensor. An N-order tensor $\mathcal{A}$ is denoted as $\mathcal{A} \in R^{I_1 \dots I_N}$ with elements $a_{i_1 \dots i_N}$. In this paper we only use 3-order tensors. All distinct Microbloggers, Contexts and Microblogs found from the database are stored in the 3-order tensor with Microblogger, the first dimension: Context, the second dimension and Microblog, the third dimension. The size of the tensor is no. of distinct Microbloggers, Contexts and Microblogs respectively. Then each record containing Microblogger, Context and Microblog of the user's posts are read from the database until all these records are exhausted. For each record the occurrence count is stored in the corresponding place of the tensor. In this way tensor $\mathcal{A}$ is constructed.

3.5 Matrix Unfolding

Given a 3-order tensor $\mathcal{A}$, three matrix unfolding operations (Symeonidis P, Nanopoulos A & Manolopoulos Y, 2010) (Drineas P & Mahoney MW, 2007) are defined in Eq. (2):

$$A_1 \in R^{I_1 x I_2 I_3} \quad A_2 \in R^{I_2 x I_1 I_3} \quad A_3 \in R^{I_3 x I_1 I_2} \quad (2)$$

where A_1, A_2 and A_3 are called the 1-mode, 2-mode and 3-mode matrix unfolding of $\mathcal{A}$, respectively. Each A_n, $1 \le n \le 3$, is called the n-mode matrix unfolding of $\mathcal{A}$ and is computed by arranging the corresponding fibers of $\mathcal{A}$ as columns of A_n (Drineas P & Mahoney MW, 2007). Matrix unfolding of Tensor $\mathcal{A}$ constructed in section 3.4 is performed for all the three modes, so that tensor is matricized as A_1, A_2 and A_3 .

3.6 Revealing Latent Associations using RSVD

Algorithm RSVD (Drineas P & Mahoney MW, 2007) (Indra R & Thangaraj M, 2019) is applied on the resultant matrices A_1, A_2 and A_3 to perform LSI. It reveals latent semantic associations that exist among Microblogger, Microblog and Context. Resultant core Tensor $\hat{\mathcal{A}}$ is the approximation of the original Tensor $\mathcal{A}$. Output of RSVD is interesting because it reveals the new associations among Microblogger, Microblog and Context. These new associations are the predictions of the future activities of the Microbloggers.

Then the whole core Tensor $\hat{\mathcal{A}}$ is stored in database with one record for each (Microblogger, Microblog, Context) triple along with the weight.

3.7 Ontology Construction

The database created in the previous section is converted into knowledge base using ontology. The classes and properties from various ontologies like SCOT[1] ontology, SIOC[2] ontology and tags[3] ontology are used in our ontology (Indra R & Thangaraj M, 2019). Each record (Microblogger, Microblog, Context and weight) of a database representing a

[1] http://rdfs.org/scot/ns#
[2] http://rdfs.org/sioc
[3] http://www.holygoat.co.uk/owl/redwood/0.1/tags/

IJASRM International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

microblog post is represented using tags:Tagging class. The entities of a tagging activity such as Microblogger is represented using class sioc:user, Context using scot:tag, Microblog using sioc:item and weight is a float connected using dataproperty with tags:tagging. The class tags:Tagging has object properties tags:taggedBy, tags:associatedTag and tags:taggedResource to link to sioc:user, scot:tag and sioc:item respectively. All microblog posts of a Microblogger are collectively represented using scot:TagCloud class. This class has two object properties viz. scot:taggingActivity and scot:contains to establish a relationship with tags:tagging and scot:tag respectively. The scot:tag is connected to sioc:user using scot:usedBy and to sioc:item using scot:tagOf. The property scot:has_tag connects sioc:item with scot:tag. If the names of the multiple Contexts in blogging posts coincide then those Contexts are aggregated to one unique scot:Tag class. The classes sioc:user and sioc:Item have other properties also to describe Microblogger and Micropost respectively. This knowledge base maintains uniqueness of all the entities, removes ambiguity of all the entities and makes faster recommendations compared to other recommendation systems

3.8 Recommendation

The constructed ontology is queried using the query language SPARQL for providing three (Microblogger, Microblog, Hashtag) recommendations. All the recommendations are based on the weight. Thus, Hashtag recommendation to users is based on their weights associated with {M, B} pair and Microblog recommendation is based on their weights associated with {M, C} pair. User recommendation for a {M, C} pair is the set of users based on their total weights which results by aggregating all Microblogs tagged by H which is in the synset(C).

4. Evaluation

The proposed RMBS architecture is implemented using Java in NetBean IDE 8.0, MySQL and MATLAB R2007b. To evaluate the proposed RMBS architecture, our own microblogging system is created. We conducted a review by 10 volunteers. Each of the volunteers, need to examine all the three recommendations. During their examination, they fill in a questionnaire about how they feel about the recommendations. They are required to give the scores for each of the recommendations. Five point [1..5] scale is used to represent the scores with 1, 2, 3, 4 and 5 representing highly unrelated, unrelated, neutral, related and highly related respectively. For example, if the volunteer feels the Hashtags are highly relevant to the blogpost in the case of Hashtag recommendation then he will give 5 points to the recommendation. Fig. 3 plots the scores for each of the recommendation by the volunteers. Above 50% of the volunteers have scored greater than or equal to 4. This result shows that, from the volunteer's judgement, the RMBS certainly provides related recommendations.

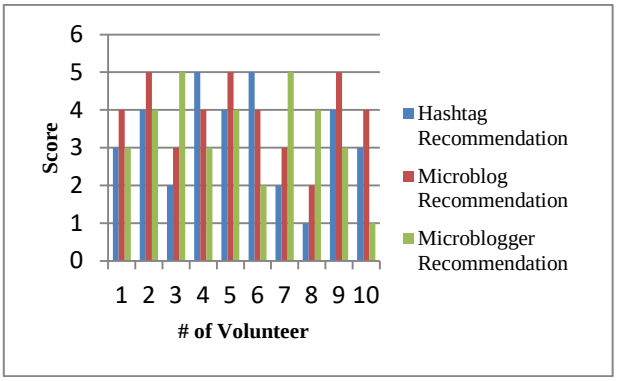

Fig. 3 The Relevance Score of the Recommendations by the Volunteers

5. Conclusion

This paper describes the proposed RMBS architecture which provides all three recommendations such as Microblogger, Microblog, Hashtag in microblogging systems. This proposed system removes the syntactic variations of the Hashtag and matches each microblog with the context. It constructs 3-order Tensor from the three entities Microblogger, Microblog, Context of the microblog post. It predicts new associations among the three entities using RSVD. Then it constructs ontology to provide recommendations using SPARQL. The main purpose of this architecture is matching microblog with the context. The prototype focused only on exploring the proposed approach and no performance aspects are performed. This work can be further enhanced by studying and analyzing its performance.

References

[1] Burak Celebi H and Uskudarli S, Content Based Microblogger Recommendation. In

Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust (SOCIALCOM-PASSAT '12). IEEE Computer Society, Washington, DC, USA, 605-610, (2012).

[2] Drineas P and Mahoney MW, A randomized algorithm for a tensor-based generalization of the singular value decomposition - Linear algebra and its applications, (2007) – Elsevier

[3] Endang Djuana, Yue Xu, Yuefeng Li, Constructing tag ontology from folksonomy based on WordNet. Proceedings of the IADIS international conference on Internet Technologies and Society, International Association for Development of the Information Society (IADIS), The East China Normal university, Shanghai, (2011).

[4] Indra R and Thangaraj M, An Integrated Recommender System using Semantic Web with Social Tagging System. International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 15, Issue 2, No. 3, (2019) In Press.

[5] Liu L, Yu S, Wei X, Ning Z. An improved Apriori–based algorithm for friends recommendation in microblog. International Journal of Communication Systems. (2018) ;31:e3453

[6] Panagiotis Symeonidis, Alexandros Nanopoulos and Yannis Manolopoulos, A Unified Framework for Providing Recommendation in Social Tagging systems Based on Terenary Semantic Analysis. IEEE Transactions on Knowledge and Data Engineering, Vol. 22. No. 2. February, (2010).

[7] Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, Yeyun Gong, Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, (2017).

[8] Qi Zhang, Yeyun Gong, Xuyang Sun, Xuanjing Huang, Time-aware Personalized Hashtag Recommendation on Social Media. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 203–212, Dublin, Ireland, August 23-29, (2014).

[9] Sana Hamdi, Alda Lopes Gancarski, Amel Bouzeghoub, Sadok Ben yahia, Enriching Ontologies from Folksonomies for eLearning: DBpedia Case. IEEE 12th International Conference on Advanced Learning Technologies (ICALT), (2012).

[10] Wenjuan Cui, Yi Du, Zhihong Shen, Yuanchun Zhou and Jianhui Li, Personalized microblog recommendation using sentimental features, IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, pp. 455-456, (2017).

[11] Yan Yu and Robin G. Qiu, Followee Recommendation in Microblog Using Matrix Factorization Model with Structural Regularization, The Scientific World Journal, (2014).

[12] Yang L., Li B., Zhou X., Kang Y., Micro-blog Friend Recommendation Algorithms Based on Content and Social Relationship. In: Yen N., Hung J. (eds) Frontier Computing. (FC 2016). Lecture Notes in Electrical Engineering, Vol. 422. Springer, Singapore, (2018)

[13] Yang Yu and Xiaojun Wan and Xinjie Zhou, User Embedding for Scholarly Microblog Recommendation. ACL, (2016).

[14] Zheng, Nan & Song, Shuangyong & Bao, Hongyun. A Temporal-Topic Model for Friend Recommendations in Chinese Microblogging Systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 45. 1-1. 10.1109/TSMC.2015.2391262, (2015).

[15] Zhuoye Ding, Qi Zhang, XuanJing Huang, Automatic Hashtag Recommendation for Microblogs using Topic-specific Translation Model, Proceedings of COLING 2012: Posters, pages 265–274, (2012).