SEMANTIC BASED E-LEARNING RECOMMENDER SYSTEM

***Nauman Sharif*, MSc, BSc (Hons) PGCE**
Institute for Knowledge Management.
Graz University of Technology, Graz, Austria
Muhammad Tanvir Afzal, PhD
Department of Computing Mohammad Ali Jinnah University
Islamabad, Pakistan
Asif Muhammad, MS,BS
Department of Computer Sciences COMSATS Islamabad, Pakistan

Abstract

Introduction of new technologies in the last few decades have brought about some innovative methods in web-based education. However many of these online courses provide universal static solution which do not cater the individual needs of the learner. Recommender system has been successfully recommending items such as books, movies, news articles etc however recommendation techniques applied in the e-learning domain are relatively new. Many of the techniques applied in the e-learning domain are generic and usually derived from other domains. This paper will present semantic based recommender system for e-learner to facilitate effective learning. We use a novel alternative to conventional recommendation techniques by considering a social network tool such as twitter which is popular for information sharing. Relevant tweets are recommended to the learner as per the current learning topic of the learner.

Keywords: E-learning, Recommender Systems, Semantic Web, personalize, twitter, Lexical matching, unigram

Introduction:

E-learning is often referred to as a mean of learning which is not restricted to physical classroom environment. Rather educational material is accessed with the help of computer and internet connection or CD/DVD anytime anywhere in the world. (n.sharif, 2012)

There are numerous LMS available in the markets which are used to deliver online courses. Examples of the most popular commercial LMS are blackboard/ WebCT, JoomlaLMS similarly free LMS are Moodle, Sakai, and Docebo.

These LMS are largely seen as a fit for all solution. A generic solution is used across different domain of education such as computer science, math, biological sciences etc. The learner is expected to interact with predefined pedagogical process which is set by the institution / instructor.

In order to discover the pedagogical aspects of WebCT Britto (2002) carried out a study. In this study instructor and learners perception of WebCT usage was analyzed. Author discovered that instructor perception of using LMS for course teaching was mainly for handiness and effective course management. On the other hand learner showed dissatisfaction in LMS as it is generic and doesn't facilitate personalize learning.

Recommendation algorithms have evolved rapidly over last ten years. Users have been benefiting from these recommendation techniques when buying movies, books and music etc. Recommender system have established its importance in such domains however recommendation techniques in e-learning domain are relatively few. Many of the techniques applied in the e-learning domain are generic and usually derived from the above mentioned domains.

The influence of internet in education takes the learners and teacher interaction into a new realm which was previously not available. Large amount of information illustrating teacher and learner interactions are continuously produced and ubiquitously available. Wide range of learning contents is available by press of a button. However this great wealth of information can also be problematic if not organized in a structured learning path. For example, a simple search on Google about a particular learning topic can generate thousands of hits and the learner has to select relevant contents manually.

This would be a daunting process for novice learner. Learner would require a few of the most relevant resources related to the task at hand rather than going through millions of generic hits.

Unfortunately traditional e-learning systems have failed to respond effectively to this new era of knowledge management. Traditional e-learning systems are seen more as course management tools to facilitate the instructor in the delivery of course contents rather than catering for the individual learns need.

There is emergence of adaptive and intelligent systems where learner can create his own learning environment which best suits his learning need rather than technology provide his learning context.

More recently personalized and intelligent e-learning systems offer personalize learning experience by constructing the learner model based on learner aims, likes and existing knowledge. A learner should be able to create, manage and organize the knowledge according to his/her personal knowledge management capabilities. Learner past learning experience and current context can be used to provide personalize and adaptive learning experience.

In this paper we proposed a model to facilitate the e-learner. E-learner will create account and complete his/her profile. System keeps history of the topics read by e-learner. Current context consists of current topic viewing, keywords related to this topic etc.

Based on the current context key terms, tweets will be fetched from twitter and stored in local database. Similarly, key terms will be extracted from learner's profile, history and current context and stored in the database.

Pre-processing will be performed on the extracted key terms to get unigrams which will be extended with the help of growbag database. These extended keywords help us to perform semantic-based matching.

Lexical matching will be performed on tweets against extended key terms. Selected tweets are sorted in descending order with respect to their occurrence frequency. Tweet are ranked by using the Relevancy Similarity. Tweets with high score finally recommended to e-learner.

Literature Review

In the literature recommendation process has been discussed by various techniques. These recommendation techniques can be classified as content-based filtering, collaborative based filtering, hybrid, Trust based and Semantic model. An e-learning recommender system can be categorize based on the above mentioned techniques.

In content-based recommendations, e-learner is recommended similar learning objects which the learner liked in the past. (Pazzani 2007)

Learner profile features are evaluated against learning objects features based on the result and new promising learning objects are recommended to the learner.

Collaborative filtering techniques on other hand provide recommendation based on either items usage history or user rating matrix. If user A and B have similar score for a group of items than it is assumed their rating score shall be same on other items as well. (Goldberg, K., 2001)

In trust base recommendation experienced learner recommendations are consider more valuable than novice learner. Each learner is assigned a trust level based on the ability of the learner and its interaction with the system.(Helic, D. 2007), (Bobadilla, J. 2009).

Existing recommender systems have the limitation of domain dependency, cold start, overspecialization and sparsity. (Marin, L 2014).

While recommendation's quality can be enhanced in various domains by combining different recommendation techniques however it should not be seen as generic solution to overcome these limitations. Users who have similar preferences in one domain may not share the same in other domains.

Semantic model can provide various advantages in personalize recommender systems. Learner's interest in particular domain can be dynamically contextualized. (Kumar, S 2015)

Next generation of recommender should consider how the personalization process can take the benefit from semantics as well as social data in order to improve the recommendation. (V. Codina 2010)

In web-based education it's possible to store most of the students learning patterns in large scale data sets and with the help of data mining techniques personalize learning profile can be created. Neil Rubens et all suggest using the artificial intelligence knowledge such as semantic filtering and recommendation systems to be used in LMSs which are geared towards eLearning 3.0. (Rubens, N., Kaplan, D., & Okamoto, T., 2014)The big data concept already exist in e-learning context as web 2.0 technologies like tweets, blogs and wikis are sharing a vast amount of learning resources on the web.

Twitter is a popular social media among student and teacher, Twitter is more open to public than Facebook and provides a fast way to exchange the ideas among peers. (Ebner, M., Lienhardt, C., Rohs, M., & Meyer, I., 2010)

An empirical study was conducted to measure the role of twitter in learning environment. Twitter utilization results demonstrate that it provides a useful mean for sharing information and collaboration among students. Students with more number of followees and followers had better grades than those students who were not actively tweeting. (Ha, I., & Kim, C., 2014).

Proposed Model

In this section we shall describe the semantic model in detail and how the semantic model can provide the useful recommendation for the e-learner. The aim is to provide the learner personalize learning activities and tasks that suits best its individual needs and as result enhance overall learning experience. Similarly recommend related tasks and activities based on previously completed tasks by learner or their peers.

A. Gold Set

In order to evaluate the effectiveness of the system we had to compare our technique with Gold Set however such Gold Set was nonexistent. We had to carry out user study to design our gold set.

We start with the ACM Computing Classification System 2012. 60 Domain Experts selected based on ACM CCS 2012. The proposed system works on research papers, so we collected 5 research papers from each expert. The Proposed system uses tweets for recommendation; therefore, we get 10 tweets per paper from domain experts. These 300 research papers with 3000 tweets will be our gold set and will be treated as bench mark. Collected research papers and tweets will be saved in local database. After performing calculations, evaluation will be performed with respect to this dataset.

B. Input for tweet ranking

Two different types of input provided to proposed system for making recommendations e.g. Paper's Metadata and tweets. Paper's metadata consists of research paper title, keywords and ACM CCS 2012 category from which it belongs. Besides metadata, complete tweets collection will also be provided.

C. Pre-processing

Before performing any kind of lexical matching, some pre-processing is required to make it read for use (Pang, C., Hendriksen, D., Dijkstra, M., van der Velde, K. J., Kuiper, J., Hillege, H. L., & Swertz, M. A., 2015). With the help of natural language processing, pre-processing will be performed on metadata contents (Dai, X., & Bikdash, M., 2015, April).

During pre-processing, word tokenization, normalization and stemming is performed. Improve greedy tokenizer used which handles digits in a smart manner. Remove Non-alphanumeric Characters from starting and ending of a string to make it meaningful. Stop words also removed during this process. Pre-processing phase resulted in the form of Unigrams.

D. Extend unigrams semantically

The unigrams have been extended with the help of two different databases. One is computer science domain specific database known as growbag (Arenas, M., Cuenca Grau, B., Evgeny, E., Marciuska, S., & Zheleznyakov, D., 2014, April) and second database is synonym based database called wordnet (Abdullah, N., & Ibrahim, R., 2015). Each unigram is compared to these databases to get domain specific and synonym based meanings of the words. These extended terms help us to perform semantic-based matching. (Ley, 2009)

E. Lexical matching

Each tweet is analyzed against extended key terms using Lexical matching. Selected tweets are sorted in descending order with respect to their occurrence frequency. Proposed system performs lexical matching on extended terms (Pang, C., Hendriksen, D., Dijkstra, M., van der Velde, K. J., Kuiper, J., Hillege, H. L., & Swertz, M. A., 2015).

F. Precision / Recall calculation

Finally precision and recall calculated on returned rows. Matched tweets are compared with the tweets given by domain experts. On the basis of matched results, precision and recall score is received (Egghe, L., 2015).

G. Evaluation

After having results for our algorithms, evaluation performed to find out the accuracy of each algorithm. Gold set used for this purpose. We compared the ranked tweets from each metadata category with the tweets given by experts. The more tweets belong to the gold set reflects the better performance of a given algorithm.

The architectural diagram of this proposed system is shown in Fig. 1.

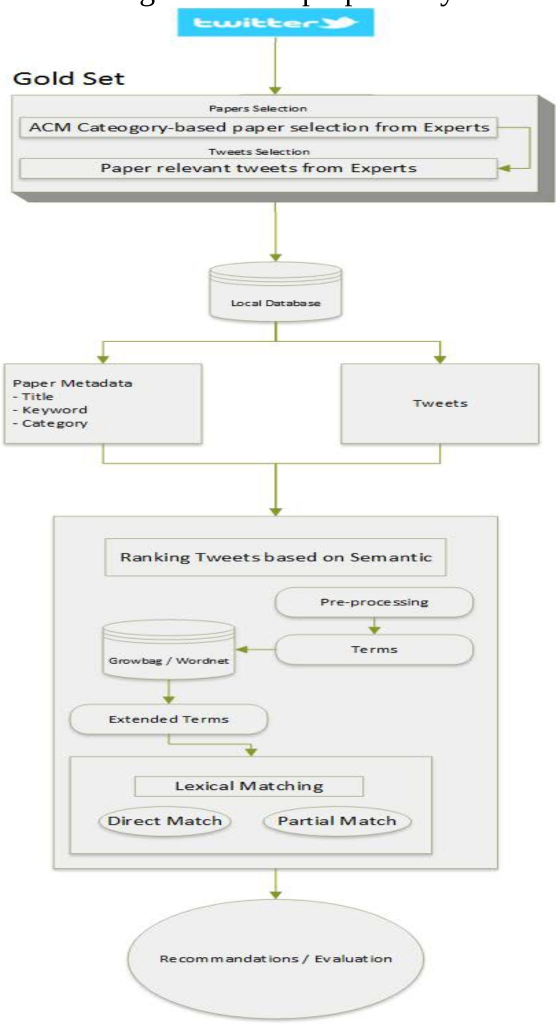

Figure 1 : Architectural diagram of proposed model

Results:

Below find graphically representation of the precision / recall score calculated in the previous step. The result is sorted in ascending order based on recall scores.

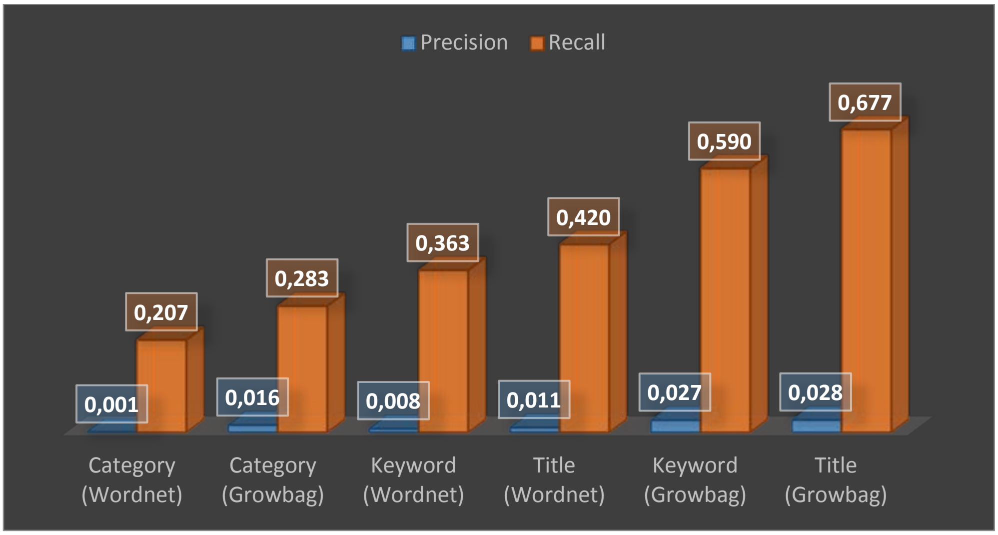

Figure 2 : Precision / Recall score of returned records

The graph shows that the recall is performing much better than precision. One reason is when we extended terms from growbag and wordnet database we received more than 10 times terms. So, retrieved rows increased many times. Hence precision went down. Recall and Precision are inversely related. An Empirical study of retrieval performance indicates that when Precision declines Recall increases. (Buckland, M. K.1994)

Another important point from the graph is Growbag extended terms performed much better than Wordnet. The reason could be that Growbag gave us computer science specific terms which are mostly used in expert's tweets. On the other hand, wordnet extended terms w.r.t. dictionary English. These terms are not too much used in expert's tweets.

Conclusion

Today's learner has the ability to utilize powerful social network tool such as twitter where the learner can independently create and redistribute contents. Hence 'digital native' learner finds general LMS structure inflexible and boring. The research community is actively engage to make the learning experience more effective with respect to the individual needs of the learner. Now the education is more centered towards the learner rather than instructor. In this paper we have presented a semantic based model to facilitate the effective learning. The system was developed and evaluated from domain experts and results were provided.

References:

Sharif, N., Afzal, M. T., & Helic, D. (2012, August). A framework for resource recommendations for learners using social bookmarking. In *Computing and Networking Technology (ICCNT), 2012 8th International Conference on* (pp. 71-76). IEEE.

Ebner, M., Lienhardt, C., Rohs, M., & Meyer, I. (2010). Microblogs in Higher Education–A chance to facilitate informal and process-oriented learning?. Computers & Education, 55(1), 92-100.

Ha, I., & Kim, C. (2014). Understanding User Behaviors in Social Networking Service for Mobile Learning: a Case Study with Twitter. Malaysian Journal of Computer Science, 27(2).

Reed, P. (2013). Hashtags and retweets: using Twitter to aid Community, Communication and Casual (informal) learning. Research in Learning Technology, 21.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World Wide Web (pp. 591-600). ACM.

Rubens, N., Kaplan, D., & Okamoto, T. (2014). E-Learning 3.0: anyone, anywhere, anytime, and AI. In New Horizons in Web Based Learning (pp. 171-180). Springer Berlin Heidelberg.

Ha, I., & Kim, C. (2014). The Research Trends and the Effectiveness of Smart Learning. International Journal of Distributed Sensor Networks, 2014.

Pazzani, Michael J., and Daniel Billsus. "Content-based recommendation systems." The adaptive web. Springer Berlin Heidelberg, 2007.

Helic, D. (2007, June). Managing collaborative learning processes in e-learning applications. In Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on (pp. 345-350). IEEE.

Bobadilla, J. E. S. U. S., Serradilla, F., & Hernando, A. (2009). Collaborative filtering adapted to recommender systems of e-learning. Knowledge-Based Systems, 22(4), 261-265.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval, 4(2), 133-151.

Marin, L., Moreno, A., & Isern, D. (2014). Automatic preference learning on numeric and multi-valued categorical attributes. *Knowledge-Based Systems*, *56*, 201-215.

Kumar, S. Manoj, K. Anusha, and K. Santhi Sree. "Semantic Web-based Recommendation: Experimental Results and Test Cases." (2015).

Victor Codina and Luigi Ceccaroni. 2010. Taking Advantage of Semantics in Recommendation Systems. René Alquézar, Antonio Moreno, and Josep Aguilar (Eds.) IOS Press, Amsterdam, The Netherlands, 163-172

Dai, X., & Bikdash, M. (2015, April). Hybrid classification for tweets related to infection with influenza. In SoutheastCon 2015 (pp. 1-5). IEEE.

Ley, M. DBLP. (2009). Some Lessons Learned. Proc. VLDB Endow. 2(2), 1493–1500.s

Pang, C., Hendriksen, D., Dijkstra, M., van der Velde, K. J., Kuiper, J., Hillege, H. L., & Swertz, M. A. (2015). BiobankConnect: software to rapidly connect data elements for pooled analysis across biobanks using ontological and lexical indexing. Journal of the American Medical Informatics Association, 22(1), 65-75.

Arenas, M., Cuenca Grau, B., Evgeny, E., Marciuska, S., & Zheleznyakov, D. (2014, April). Towards semantic faceted search. In Proceedings of the companion publication of the 23rd international conference on World Wide Web companion (pp. 219-220). International World Wide Web Conferences Steering Committee.

Abdullah, N., & Ibrahim, R. (2015). Managing Information by Utilizing WordNet as the Database for Semantic Search Engine. International Journal of Software Engineering and Its Applications, 9(5), 193-204.

Egghe, L. (2015). Quantitative analysis of the Recall-Precision relationship in information retrieval.

Buckland, M. K., & Gey, F. C. (1994). The relationship between recall and precision. JASIS, 45(1), 12-19.