

Recommendation System for Mining Financial News from Web

Bhumika D Karwa¹, Krunal Panchal²

¹Student, Department of Computer Engineering, L. J. I. E. T, Ahmedabad, Gujarat, India

²Assistant Professor, Department of Computer Engineering, L. J. I. E. T, Ahmedabad, Gujarat, India

ABSTRACT

Now-a-days people tend to read news on line instead of reading via physical newspaper subscription. Meanwhile, finance plays a more and more crucial role in the world. The critical problem people are confronted with is the volumes of financial news are overwhelming to the users. To deal with this challenge, we propose a Action based recommendation technique which help users find the articles that are interesting to read. Action Based Recommendation providing real time dynamic recommendation of genuine news to all the visitors of the website irrespective of been registered or unregistered. To settle the ambiguity problem, ontology is employed to represent the unstructured text data in the form of key concepts which are all stored in the financial domain ontology. Finally, the experiment conducted on a financial news dataset demonstrates that the proposed technique significantly outperforms the performance of a traditional recommender.

Keywords— Financial News, Recommendation, Ontology, Weblog, Registered user, unregistered user

1. INTRODUCTION

With the rapid development of World Wide Web, people tend to read news on line instead of reading via physical newspaper subscription. Meanwhile, finance plays a more and more crucial role in the world. The critical problem people are confronted with is the information overload which means the volumes of financial news are overwhelming to the users. One possible solution to deal with the news items overload problem is the use of recommender systems. Typically such recommenders employ a user profile and aim to recommend news items that best match this user profile.

There are currently three different ways on how a news recommender can recommend news articles: a content-based, a collaborative filtering, and a hybrid recommendation. Content-based recommendation focuses on existing user preferences and searches for news items which are similar to the user preferences. The collaborative recommendation focuses on other users which are similar to the person browsing for news items and what articles they found interesting. The hybrid method is a mix of the two previously mentioned recommendation systems and tries to combine the better of both. Some argue that within content based recommending systems one could distinguish between two subtypes: a recommendation system that simply takes every word into account (without paying any attention to word meanings) on the one hand and a semantic approach on the other hand, which tries to capture the meaning or sense of each word.

Content-based methods play a central part in recommender systems, as it is able to recommend information that has not been rated before and accommodates the individual differences between users. One of the disadvantages of it is the lack of semantics. To deal with the issue, we use ontology to store lexicalized financial news concepts, relations and synsets.

This paper focuses on providing real time recommendation to online users who can be either registered or unregistered. Two different approaches are proposed to provide effective Action Based recommendation. In news based technique, recommendation is provided to unregistered user based on IP address as obtained from log file. Another technique is user based technique which provides recommendation to registered customer based on session constructed for each unique user based on users' navigation.

The paper is organized as follows: Section II is dedicated for the related work. Section III describes problem definition. Section IV shows proposed work. Section V describe experiment result and analysis. Section VI describes conclusion and ends with references.

2. RELATED WORK

Wouter IJntema et al. (2010)^[8] developed a recommendation system, Athena, to provide ontology-based recommendation for the news feed system. It extends the Hermes framework, a framework used to build a news personalization service, with the help of ontology to determine the semantic relations between terms and concepts. It uses ontology to store concepts and their relationships to the news items.

Jiahui Liu et al(2010)^[11] describes a research on developing a personalized news recommendation system based on profiles learned from user activity in Google News. Based on the analysis of Google News users click logs, Bayesian framework is being developed for predicting users' current news interests from the activities of that particular user and the news trends demonstrated in the activity of all users. By combining the content-based recommendation mechanism which uses learned user profiles with an existing collaborative filtering mechanism personalized news recommendations is being generated.

Lingling Zhang et al(2015)^[2] presented OF-IDF method to represent the unstructured text data in the form of key concepts, synonyms and synsets which are all stored in the domain ontology. For users, the recommendation algorithm builds the profiles based on their behaviours to detect the genuine interests and predict current interests automatically and in real time by applying the thinking of relevance feedback.

Prajyoti Lopes et al(2015)^[12] focuses on providing real time recommendation to online users who can be either registered or unregistered. This technique makes use of traditional web usage mining steps for data acquisition and data cleaning and finally to construct useful session. Two different approaches are proposed to provide effective recommendation. In product based technique, recommendation is provided to unregistered user based on IP address as obtained from log file. Another technique is user based technique which provides recommendation to registered customer based on session constructed for each unique user based on users' navigation. Results show that both the techniques provide better recommendation quality and accuracy.

Ricardo (2004)^[13] uses a Bayesian classifier to select articles of interest to a specific user, according to his profile. The articles are extracted from web pages and displayed in a zoomable interface-based browser on a PDA. Interests may change over time, making it important to keep the profile up to date^[13]. The system monitors the users' reading behaviors, from which it infers their interest in particular articles and updates the profile accordingly.

Kang and Choi et al(2011)^[6] proposed an ontology based Recommendation system in which the ontology is used to encode the long term and short term preference information. The user preference ontology is constructed from the concepts of the general domain ontology together with the documents that the user visited. Recommendation is made based on the similarity between ontological concepts and terms.

3. PROBLEM DESCRIPTION

In Existing System, the recommendation algorithm build the profiles based on their behaviors to detect the genuine interests and predict current interest by applying the thinking of relevance feedback. But every time user interaction with the system is not possible. To overcome this issue we proposed Action Based Recommendation System. Action Based Recommendation providing real time dynamic recommendation of genuine news and predict current interest automatically to all the visitors of the website irrespective of been registered or unregistered.

4. PROPOSED WORK

The contribution of this paper is two-fold. Firstly, we construct a specific financial domain ontology which try to store all the paramount information about the certain amount of financial news. Secondly, this paper proposes a new method called Action based Recommendation System providing real time recommendation to online users who can be either registered or unregistered to represent the unstructured text data.

Financial Domain Ontology

Financial domain ontology try to store all the paramount information about the financial news. The ontology is a shared and common understanding of some domain that can be communicated across people and computers. The domain ontology includes the key concepts that capture the semantic context of the articles as well as the relations between them such as hasCompetitor and isCEOof. For example, if a user is interested in the news of Google, it is likely that he is fond of the information of Yahoo. Moreover, for every concept in the ontology, a synonym property is defined to eliminate the ambiguity with lexical representations of a concept. Recommenders that focus on the ontology might produce faster and more accurate recommendations than the term-based recommenders, since they don't need to consider all words.

Action Based Recommendation Technique

This technique provides providing real time recommendation to online users who can be either registered or unregistered. This technique makes use of traditional web usage mining steps for data acquisition and data cleaning and finally to construct useful session. Two different approaches are proposed to provide effective recommendation. In news based technique, recommendation is provided to unregistered user based on IP address as obtained from log file. This method has an edge over traditional techniques which provide recommendation to unregistered user based on cache memory. Another technique is user based technique which provides recommendation to registered customer based on session constructed for each unique user based on users' navigation.

4.1 Proposed Algorithm

INPUT:

Web News Repository and Web Log files of registered and unregistered user

OUTPUT:

Recommendation List for registered and unregistered user

ALGORITHM:

1. Start
2. Collect News Web Repository and Web Log File for user history
3. Preprocessing of Web Repository and Web log file
4. Structured Representation of Financial News using Financial Ontology
5. Use Action Based Recommendation Technique
6. If the user is registered
7. News Recommendation List using recent UserID in log file
8. If the user is Unregistered
9. News Recommendation List using recent IP addresses in log file
10. Stop

4.2 Proposed System Flow-Diagram

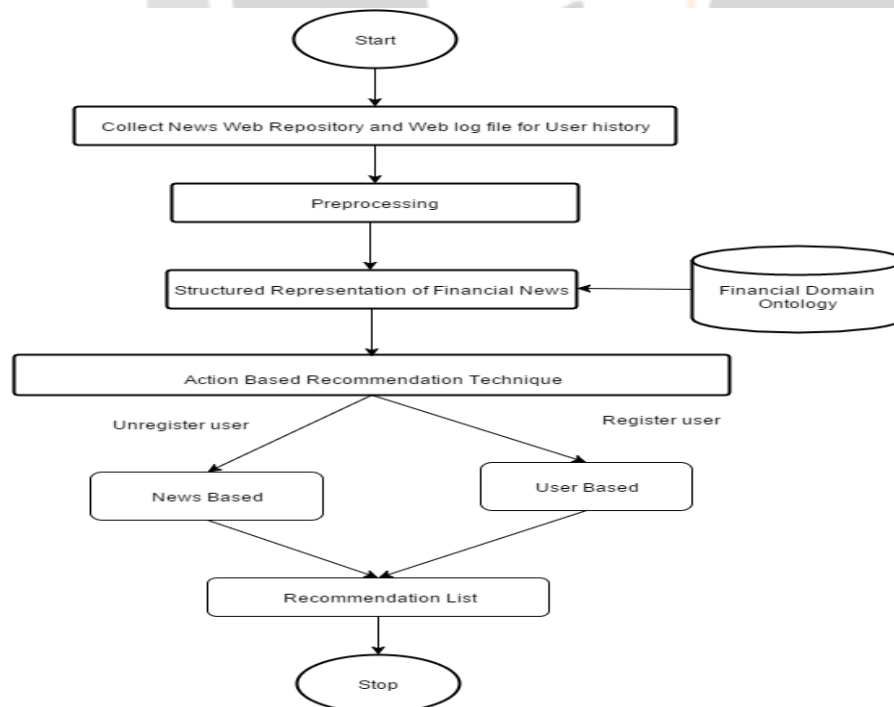


Figure 2: Proposed System Flow Diagram

1. Data Collection

The proposed flowchart starts from collecting Financial News Web Repository and entire clickstream data and maintained in a log file. Different techniques are then applied on cleaned log file to provide effective personalized recommendation.

2. Data Preprocessing

For effective data analysis, good and better quality of data should be served as an input. The collected web log data consists of lot of irrelevant and inconsistent data and needs to be cleaned for effective mining.

A. Web Log File Preprocessing

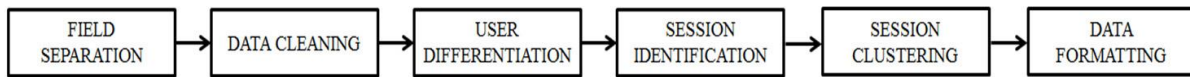


Figure 4.2. Steps for Web Log File Pre-processing

1. **Field separation:** It focuses on separating individual fields by making use of separator character such as space.
2. **Data cleaning:** Data cleaning is a process of filtering out irrelevant and outliers' data. It eliminates all irrelevant items by checking the suffix of the URL name. Therefore, all log entries with filename suffixes such as gif, jpeg, GIF, JPEG, and JPG are removed. Data cleaning reduces the total number of records and also log file size.
3. **User differentiation:** It is important to distinguish between different users for analyzing different user access behavior patterns. A different user ID will be assigned to different IP address. In case of same IP address referrer information and browser details will be used to distinguish among different web users.
4. **Session identification:** A session is defined as an ordered sequence of web pages visited by a user. A new session is constructed based on new IP address. Each new IP addresses implies (correspond) to an unique user. A maximum session time limit is considered to be 30 minutes.
5. **Data formatting:** Finally data will be formatted to appropriate tabular format for further analysis.

B. Web Repository Preprocessing



Figure 4.3. Steps for Web Log File Pre-processing

1. Remove all HTML Tags using regular expressions.
2. Remove StopWords from the news data. Usually these words are filtered out from search queries because they return vast amount of unnecessary information such as the, is, at, which, and on.
3. After removing stopwords, stemming of words into their root form. For example "fishing", "fished", and "fisher" to the root word, "fish".
4. Financial Ontology is employed to represent the unstructured text data in the form of key concepts which are all stored in the domain ontology.

5.Action Based Recommendation System suggests two different methodologies namely product based and user based recommendation system. Both the technique generates a list of recommended products to the individual users for providing personalized recommendations.

a) News based recommendation technique

This technique is more suitable to provide recommendation for unregistered users. The beauty of this technique is to provide effective recommendation even in the case, where the user clears the cache memory on his/her browser. Also it tries to provide better recommendations if different users access the same system and browser by providing a combination of recommendation based on most recent session and timestamp.

b) User based recommendation technique

This technique is more suitable to provide recommendation for registered users. Based on user's navigational detail appropriate recommendation is provided.

For Unregistered user IP address is the deciding factor and for registered user session id .Both techniques use recent session timestamp.

5. EXPERIMENT RESULT AND ANALYSIS

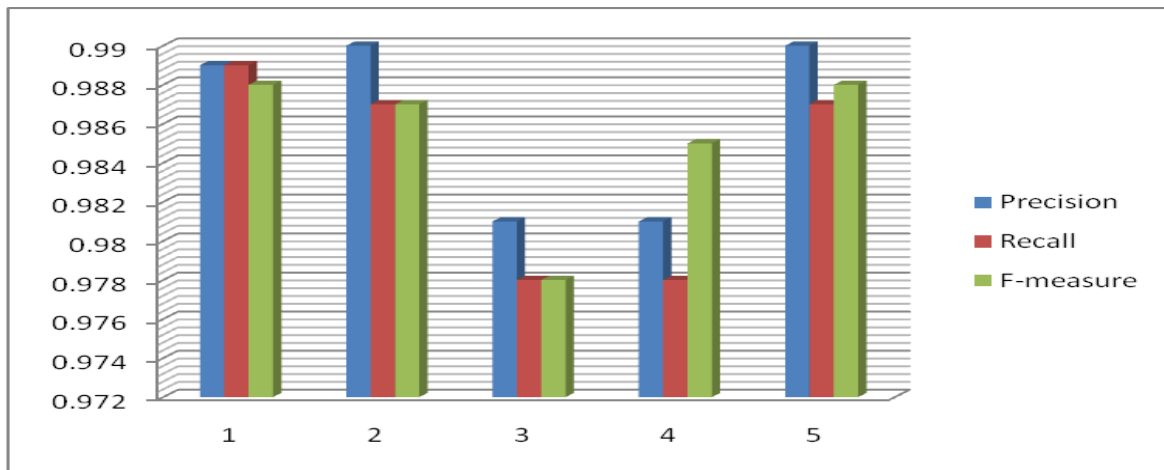


Chart 1 :-Precision, Recall and F-measure of recommendation list

According to the foregoing experimental results, the algorithm we proposed outperforms the baseline method at precision, recall and F-Score. It indicates that the financial ontology can characterize user’s reading habit more precisely.

Table 1: Runtime Analysis

No. of News	500	1000	2000
Runtime(s) without Ontology	0.002	5.433	20.677
Runtime(s) with Ontology in Baseline	0.428	0.937	1.897
Runtime(s) with ontology in Our method	0.312	0.786	1.378

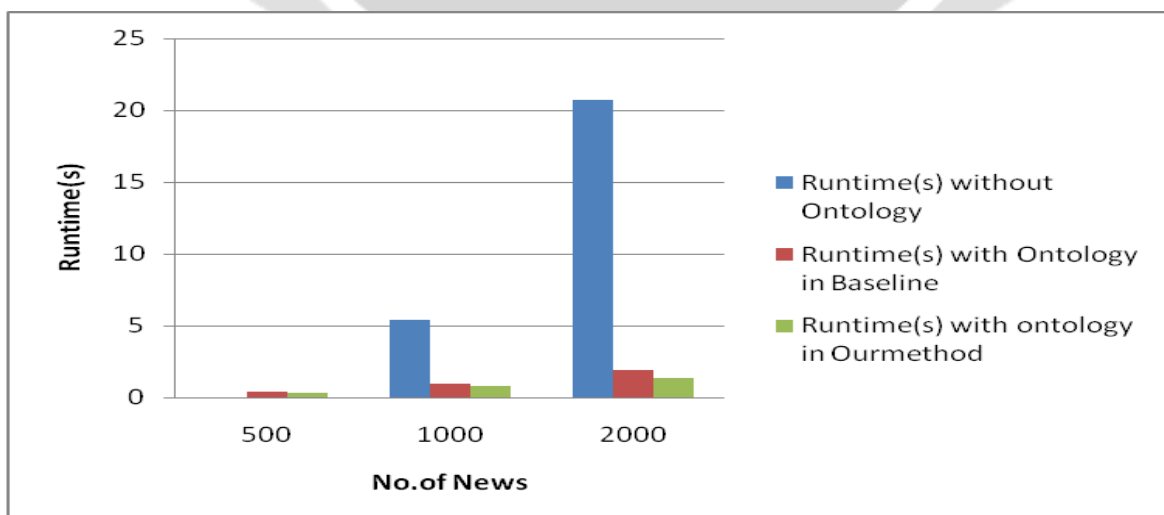


Chart 2 :- Runtime(s) Analysis

6. CONCLUSIONS

In this paper, we propose a financial news recommendation approach which aims to deliver the interesting news articles to the users. To settle the ambiguity problem, we first constructed a specific financial domain ontology which stores key concepts and relationship between them. In the second place, Action Based Recommendation technique provide real time dynamic recommendation of genuine news to all the visitors of the website irrespective of been registered or unregistered. Finally, we conducted an experiment to compare the new approach with the baseline method which demonstrated that the technique we proposed markedly improved the quality of news recommendation. In future work proposed approach also recommends current trend news to all users. However it is important to evaluate our technique with other recommendation approaches and check its effectiveness.

7. REFERENCES

- [1] Liu J, Dolan P, Pedersen E R. Personalized news recommendation based on click behaviour. Proceedings of the 15th international conference on Intelligent user interfaces. ACM, 2010: 31-40.
- [2] Rui Ren, Lingling Zhang, Limeng Cui, Bo Deng, Yong Shi, "Personalized Financial News Recommendation Algorithm Based on Ontology" in *Procedia Computer Science* 55 (2015) 843 – 85
- [3] Goossen F, IJntema W, Frasinca F, et al. News personalization using the CF-IDF semantic recommender. Proceedings of the International Conference on Web Intelligence, Mining and Semantics. ACM, 2011: 10.
- [4] Capelle, M., Moerland, M., Frasinca, F., Hogenboom, F. Semantics-Based News Recommendation. In: Akerkar, R., B_adic_a, C., Dan Burdescu, D. (eds.) 2nd International Conference on We Intelligence, Mining and Semantics (WIMS 2012). ACM (2012)
- [5] F. Frasinca, J. Borsje, and L. Levering. A Semantic Web-Based Approach for Building Personalized News Services. *International Journal of E-Business Research*, 5(3):35-53, 2009.
- [6] Fikadu Gemechu, Zhang Yu, Liu Ting. "A Framework for Personalized Information Retrieval Model", Proceedings of IEEE transaction, 2010
- [7] K. G. Saranya and G. S. Sadhasivam, "A personalized online news recommendation system", *International Journal of Computer Applications*, pp 6–14, November 2012
- [8] IJntema W, Goossen F, Frasinca F, et al. Ontology-based news recommendation. Proceedings of the 2010 EDBT/ICDT Workshops. ACM, 2010: 16.
- [9] Adomavicius, G., Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering* 2005,17(6):734–749.
- [10] Chen, M., Han, J., and Yu, P. (1996) *Data Mining: An Overview from a Database Perspective*. *IEEE Transactions on Knowledge and Data Engineering*. 8, 866-883.
- [11] Eirinaki, M. and Vazirgiannis, M. (2003), "Web mining for web personalization", *ACM Transactions on Internet Technology* 3(1), 1–27
- [12] Prajyoti Lopes and G Bidisha Roy, "Dynamic Recommendation System Using Web Usage Mining for E-commerce Users", *International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Elsevier 2015, IEBN:978-1-4503-0915-8, pp 60-69.
- [13] Carreira, R., Crato, J. M., Gonçalves, D., Jorge, J. A. Evaluating adaptive user profiles for news classification, Proceedings of the 9th international conference on Intelligent user interfaces, ACM(2004), ISSN: 0986 – 8457 ,pp 8–14.