Recommending Sources in News Recommender Systems

Özlem Özgöbek[1 2], Jon Atle Gulla[1] and R. Cenk Erdur[2]
[1]*Department of Computer and Information Science, NTNU, Trondheim, Norway*
[2]*Department of Computer Engineering, Ege University, Izmir, Turkey*

Keywords: Recommender Systems, News Source, News Recommendation.

Abstract: Recommender systems aim to deliver the most suitable item to the user without the manual effort of the user. It is possible to see the applications of recommender systems in a lot of different domains like music, movies, shopping and news. Recommender system development have many challenges. But the dynamic and diverse environment of news domain makes news recommender systems a little bit more challenging than other domains. During the recommendation process of news articles, personalization and analysis of news content plays an important role. But beyond recommending the articles itself, we think that where the news come from is also very important. Different news sources have their own style, view and way of expression and they may give the user a complete, balanced and wide perspective of news stories. In this paper we explain the need for including news sources in news recommendation and propose a news source recommendation method by finding out the implicit relations and similarities between news sources by using semantics and association rules.

1 INTRODUCTION

News recommendation is a challenging task which includes many difficulties compared to other recommendation domains. News domain has a very dynamic environment with usually hundreds of new articles published every hour. While the number of online articles increases, the recency and popularity of articles change too fast which makes the recommendation more challenging. In (Ozgobek et al., 2014) challenges in recommender systems and news recommenders are explained in detail. To be able to make suitable recommendations to the user, the recommender system needs the detailed user information and/or the content of news items.

News recommendation is considered as the news article recommendation. So each item that the recommender system recommends is a news article. There is a vast amount of research going on about news recommender systems and there are very successful results. But all the news recommender system research is too much focused on the articles, analysis of the text, finding similarities between news articles and predicting user's like on each article. There is no news recommender system which takes the source of the news into account that we could find.

Different news sources may have different specializations like sports, science etc. Even though they publish the same story with other sources the way they express the news, the words they select to use may be different. Each news source has its own style, view and way of expression that some users like and some does not. So it is important to consider the differences between news sources and to recommend articles from the sources that users like.

Serendipity problem for recommender systems addresses the problem of recommending similar or the same items with the already recommended ones. A different article describing the same event should not be recommended while keeping the diversity of recommendations (Lops et al., 2011). For news recommender systems it is challenging to solve this problem because it is harder to understand the differences of the meaning of the whole article. It is really hard to distinguish the news articles if they are the same news written in a different way or different news stories related to the same topic. For example two news articles may be identical which are telling the same story or they may be the different parts of one story. By recommending news sources we mitigate this problem. When there is more than one news article on a specific topic coming from the same news source, it is usually possible to say that they are different news articles on the same topic. We can recommend very similar articles from the same source without worrying about serendipity problem.

Özgöbek Ö., Gulla J. and Erdur R..
Recommending Sources in News Recommender Systems.
DOI: 10.5220/0005489205260532
In *Proceedings of the 11th International Conference on Web Information Systems and Technologies* (WEBIST-2015), pages 526-532
ISBN: 978-989-758-106-9

In (Ofcom, 2013) it is stated that less than half of online news users (45%) use only one source to get the news. Similarly in (Media, 2012) it can be seen that the average number of online news sources consumed changes between 1.4 and 2.4 according to the demographic differences, where the total average is 2.0, including websites and apps.

News recommender systems usually include news resources that are hard coded by the developers. Even though it is possible to give the right to select the news sources to the user, the user would select the most trusted, liked or known news sources for herself. This prevents the users to discover new sources of news articles and narrows the scope of the news recommender system. Since many online news readers usually check only a few online news sources in their daily routine, it is hard for users to discover new sources and gain a wider perspective by what they read. In (Media, 2012) it is stated that "Our qualitative research showed that reasons for multi-sourcing can often be active choices, where a person seeks to get a balanced, complete view of a story from across a range of providers and platforms.". So beyond making personalized news article recommendation, it is also important to make news source recommendations and consider the news sources in article recommendations. Among many online news sources it is also a challenge to find the source that the users would like. Considering the news sources including websites, blogs and other possible sources during the news recommendation process, helps us to increase the quality of recommendations, provides better solutions for dealing with the news recommender system challenges and gives the chance to the users to discover different sources which they may like.

In this paper we propose a news source recommendation method. To be able to recommend news sources, we used two different methods to find out the implicit relations and similarities between news sources. The first method is a semantic analysis of news content. For this method we built a small scale ontology which includes important terms from specific news sources. The second method is the association analysis which is a data mining method. We used the association analysis to find out the relations between news sources according to the users' reading patterns. Then we compared and discussed these two different approaches to find the hidden relations between news sources.

The rest of the paper organized as follows. In Section 2, we give information about the background work done within the SmartMedia project which this work also belongs to. Section 3 describes the related work about semantic news recommender systems and association rules for recommender systems. The details of our approach is explained in Section 4. In Section 5 the results and discussion is explained. Finally in Section 6 conclusion and future work is given.

2 BACKGROUND WORK

SmartMedia project [1] in Norwegian University of Science and Technology (NTNU) was started in 2011 in close collaboration with Scandinavian media industry. With this project it is aimed to present the online news information in an effective and personalized way to the users while considering the point of view of the journalists and media companies. The main focus of the project is recommender systems and semantic search.

Within the SmartMedia project it is presented a user profiling approach for the mobile news recommender systems (Gulla et al., 2013). In this work the users' actions on the mobile device is observed and by using this information an approach for learning the user profiles is proposed. The implemented mobile news recommender of SmartMedia project is proposed in (Tavakolifard et al., 2013). There is also ongoing work about the multi-platform implementation of our news recommender system. The progress of building a complete and publicly available dataset in Norwegian news domain is continuing (Ozgöbek et al.,) within the SmartMedia project.

The proposed work in this paper is also continuing as a part of the SmartMedia project.

3 RELATED WORK

3.1 Semantic News Recommenders

Semantic approach is one of the methods for recommendation. The main motivation to use semantics in recommender systems is to be able to use the cultural and linguistic background knowledge of the content (Peis et al., 2008). The use of semantics reduces ambiguity compared to keyword based systems, it allows hierarchical representation of concepts and inference (Cantador and Castells, 2009). In (Lops et al., 2011), the semantic recommenders are grouped according to their use of different semantic approaches. Since the challenges to solve and approaches applied to solve these challenges differs from domain to domain (Ozgobek et al., 2014), it is also possible to see different semantic approaches for different domains. Although

[1] http://research.idi.ntnu.no/SmartMedia/

there are a lot of semantic recommender system research available, in this section we are going to consider the semantic news recommenders which is our main focus of interest.

In (Cantador and Castells, 2009), it is proposed a news recommender system News@Hand, which uses semantic technologies to provide recommendations. Ontolgies are populated from the news contents by extracting the noun terms. Also Wikipedia articles are used to populate ontology classes. To overcome the data sparsity problem in user profiles, it is proposed a mechanism to expand the user preferences. The current context is also considered and defined in a way that the importance of concepts decreases within time for making better context aware recommendations.

Hermes framework uses a semantic approach to build personalized news service. (IJntema et al., 2010) It is proposed an extension to the Hermes framework for the semantic news recommendation which is called Athena. For the recommendations, first a user profile is constructed by using the user's reading history. Then by using the different similarity measures and the ontology populated by the Hermes framework recommendations are done. The news recommendation method which is implemented in Athena is proposed in (Goossen et al., 2011). In this work, the well known method TF-IDF (Term Frequency - Inverse Document Frequency) for content based recommenders is applied to the semantic recommenders as CF-IDF (Concept Frequency - Inverse Document Frequency). CF-IDF considers only the key concepts in the news articles where TF-IDF considers all the terms. Similarly, in (Capelle et al., 2012) it is proposed two new methods called Synset Frequency - Inverse Document Frequency (SF-IDF) which uses the WordNet synonym sets and Semantic Similarity (SS) to calculate the similarities between news items. It is used to recommend news items based on the user behavior profile and semantic similarity measure. The proposed methods are implemented as the Ceryx framework which is an extension of Athena.

(Rao et al., 2013) proposes an ontology based similarity model to calculate the news-user similarity in a semantic news recommender system. The ontologies are populated by using the online encyclopedias as DBPedia. The proposed similarity model is built on these ontologies and the background information is used to measure the similarities between news articles and users.

In (Lašek and Vojtáš, 2011) it is proposed a semantic information filtering workflow in the news filtering use case. The workflow includes many steps including entity identification, semantic data crawling and building user profile. This work aims to use the advantages of semantic background information used together with the user profiles and improve the results of name entity recognition by the help of user feedback.

Even though there are many works on personalized news recommendation, there is no research which considers the news sources that we could find. We think that having a wider perspective of news stories around the world or ignoring the news sources that the user does not want to read is important. So considering news sources is an important aspect of news recommender systems to work on.

3.2 Association Rules for Recommender Systems

Association analysis is a data mining method which is used to discover the hidden relationships of items in large datasets (Pang-Ning et al., 2006). The relations are represented as association rules which is shown as $X \rightarrow Y$ where X and Y are disjoint itemsets. The strength of association rules is measured with two metrics called support and confidence. Support defines the proportion of the number of transactions containing X and Y together to the total number of itemsets in the dataset. Support is shown as $s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$. Confidence defines the proportion of the transactions which includes X and Y together to the number of transactions only contains X. Confidence is shown as $c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$.

Association analysis is widely used in many areas for different purposes. A very common and successful usage of association analysis is for the supermarkets to find out the hidden relations between products and change the sales policies accordingly. By analyzing the users' buying patterns it is possible to find a relation between two disparate products (Pang-Ning et al., 2006). There are also some studies about the association rule mining in recommender systems. In (Mobasher et al., 2001) a personalization framework based on association rule discovery is proposed. Association rules are also used to develop more secure recommender systems. For example in (Sandvig et al., 2007) an association rule based algorithm is presented to prevent the profile injection attacks in recommender systems. In (Lemdani et al., 2010) it is presented a collaborative filtering method by using the association rules. Similarly (Sun et al., 2005) proposes a method which applies quantitative association rules in collaborative filtering and compares with the conventional Pearson method.

4 OUR APPROACH

In our approach to the news source recommendation, we used two different methods to find the similarities between news sources. Our aim was to find out some rules between news sources according to the users' reading patterns. We preferred to use association rules because it gives exactly what we are looking for. By using association rules we discovered some rules like: "Who reads from source X also reads from source Y." which can be shown as X → Y. This approach is a well known and used method in recommender systems by itself. In addition to the findings from association rules we wanted to see if the source X and source Y are semantically related. To do that we built a small ontology to detect the similarities between news sources. The details and results are given below.

4.1 Dataset

To test our approach we are using the YOW dataset (Wolfe and Zhang, 2010). YOW dataset includes two parts. The first part is publicly available on the web [2] and it is collected in a user study at Carnegie Mellon University in 2004. This part of the dataset contains the article information (document id, source etc.) read by users, explicit feedback (user likes) and very detailed implicit user feedback like the time spent on each article. A small example from the dataset is shown in Figure 1. In the dataset, each user, article and source has a unique id. In Figure 1, we can see which user (user id) read which article (DOC ID), how much she liked (user like) and the source number of the news article (RSS ID). YOW dataset contains data collected from 25 users in total. The number of read articles from each source by each user included in this part of the dataset is used to find out the possible similarities between news sources.

user_id	user_like	RSS_ID	DOC_ID	TimeOnPage
80	1	5	504	1842
91	2	2970	5281	4046
56	4	6900	7906	80466
63	5	6900	7906	42181
63	5	6900	7908	69239
73	4	6900	7908	365047
76	4	6900	7908	18516
59	3	6900	7909	40172
63	1	6900	7909	36693
59	2	6900	7912	31813

Figure 1: A small example of the dataset.

The second part of the dataset is crawled from RSS feeds (Zhang, 2005). This part includes textual information about news articles including the headline, an introductory text, URL and source number (RSS ID). So by using the RSS IDs we can group the news articles according to their sources. Since there are many news articles from each source, it is possible to analyse the textual information for specific sources. By using text analysis we extracted the key concepts for every news source and built an OWL ontology for the ontology based similarity detection of news sources.

4.2 Ontology based Similarity Detection of News Sources

Ontologies and semantic reasoning is a useful way of revealing relations between entities. For the news sources, each source may have different areas of focus like sports, politics, movies etc. Or they may use different terminology for news stories. On one hand, finding important terms used in articles from a news source can tell us more about the properties of that source. On the other hand, finding common similar terms of different news sources gives us the chance to discover the similarities between news sources. The YOW dataset contains the headlines and short introductory text of each article. So we started our progress of revealing the similarities between different news sources using ontologies by finding the important terms of articles contained in each news source.

For ontology construction we use Protégé [3] and TerMine [4]. Protégé is a widely used open source platform to build ontologies. TerMine is a text mining tool to extract candidate terms from a text. It can also be used as a Protégé plug-in.

The class hierarchy in the ontology is built according to the common news article categorization. So it is easy to see the detailed category and topic similarities between news sources. A small example of the ontology class hierarchy is shown in Figure 2. Each individual in the ontology has an object property 'has source' which shows the source(s) of entites.

The developed ontology is a very small scale ontology only to study the possibility of recommending news sources by using semantic relations.

When we analyse the resulant ontology which contains more than 350 individuals from 26 different sources and 16 main categories, we found some relations between news sources. In Figure 3 relation between two news sources number 8156 and 8174 is shown as an exmple. They both include common topics about actors. So it is possible to say that these two

[2] http://users.soe.ucsc.edu/ yiz/papers/data/YOWStudy/

[3] http://protege.stanford.edu/

[4] http://www.nactem.ac.uk/software/termine/

Figure 2: A small example of the ontology class hierarcy.

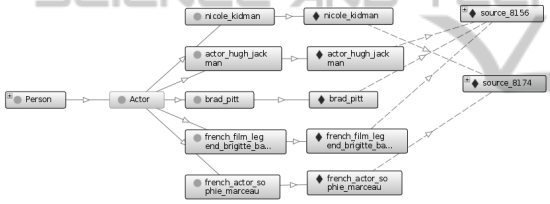

Figure 3: Relation between two news sources as an example.

news sources include related topics and if one user likes one of the sources, it is possible to recommend her articles from the other news source.

Also there is a possibility that even though the sources are related in topic based, the user might not like one of them because of the different language used in that source or the different view of the writers over the topic (how they discuss on the topic). This aspect of the news source recommendation will be handled in future work. In this work our aim is to show that it is possible to make recommendations based on semantic analysis and machine learning techniques together.

In our analysis of news sources like the example shown in Figure 3, we consider a source relation only if that source has two or more common entities with the other sources. By doing this we eliminate the chance factor on discovering relations of news sources. For example, the name of actor might be included in a news article related with sports, but the news source can be a source related with only sports, not with movies or actors. We eliminate the chance factor in first step when we extract the important terms from the text contained in a news source. And by choosing only the relations containing more than one common entity we make sure that we find the related news sources.

In Table 1 it is provided some of the common relations of news sources discovered by the semantic relations in the ontology. In this table X-axis denotes different news sources and y-axis shows different categories. The numbers in the table denotes the frequencies of instances for each source in different categories.

4.3 Association Rule based Relation Detection of News Sources

A user's interest to a specific news source can be understood by looking at the number of articles read from that source. For a user, as the number of read articles from a specific source increases, it is more likely that the user is interested in that news source. If a user is interested in a few news sources, these sources may be similar to each other or they may have similar impact levels on readers. So finding the underlying relations of news sources that is not clearly visible becomes an important issue on news recommendation. In this work we used association analysis to find out the implicit relations between news sources according to the users' interests.

To analyze the users' reading patterns by association rules, we extracted the number of reads per user from each news source for 24 users and 26 news sources. In Table 2 a small example of users' number of reads from different sources is shown. The news sources used in this experiment are the same news sources that we used to build an ontology which is described in previous section. For applying the association analysis we used Orange [5] which is an open source data visualization and analysis tool.

In the analysis of the results to prevent considering the rules which may occurred by chance, we ignore the association rules which has a support number below 0,4 and confidence number below 0,7. Especially the support is known as an important measure for the rules occur by chance (Pang-Ning et al., 2006).

5 RESULTS AND DISCUSSION

In this work we proposed a news source recommendation method which uses two methods to find the

[5]http://orange.biolab.si/

Table 1: Some relations of news sources extracted from ontology. X-axis denotes different news sources.

	1265	1302	1892	3616	8152	8155	8156	8157	8172	8174	8185	8195
Business	3					8			6			
Entertainment					3	5	8			3	2	2
Economy						5						
Event	2						6			3		2
Health		3	5					8				
City				3						2		2
Sports												
Software	6								2			
Web					3				8			

Table 2: Users' number of reads from different sources.

	Source X	Source Y	Source Z	...
User A	35	2	15	...
User B	28	73	6	...
User C	4	12	26	...
...	...	...	...	...

Table 3: Ontology and association rule results.

Source 1	Source 2	Support	Confidence
8156	8195	0,619	1,000
8153	8195	0,572	1,000
8156	8155	0,461	0,769
8172	8155	0,429	0,900
8172	8152	0,333	0,700

most relevant news sources. The first method contains semantic analysis of news content from different sources. As it is explained in detail in Section 5.2 we built an ontology where each entity belongs to a news source. So by looking at the number of entities within a class which belongs to a specific news source it is possible to see the topics included in that news source. The second method uses the association analysis to find out the relations between news sources according to the users' reading patterns. When we merge the results from these two methods we get the most relevant news sources that the system can recommend.

Our results show that it is possible to find a correlation between the related news sources and users' reading patterns. If there is an extracted rule like X $\rightarrow$ Y also there is usually a semantic relation between news sources X and Y. In Table 3 it is shown the top 5 results of this correlation between association rules and the ontology. All the relations seen in this table occurs both in ontology and association rules. Each source is represented by a four digit number, also the association rule support and confidence values are given.

We see the news source recommendation in three different aspects:

- Possible solution to the serendipity problem.
- Discovery of news sources for users. It gives a wider perspective of news stories.
- Better personalized news recommendations by considering news sources. Some people like to read from several different sources, some like to follow only one.

6 CONCLUSION AND FUTURE WORK

In this paper we presented a news source recommendation method. As the number of online news sources increases it becomes harder for users to find the suitable news sources for themselves. A user may spend many hours to discover new sources of news which she likes. It is also possible that she may never find some sources that she would like. The same news topic may be represented differently in different news sources. So when the user likes how news topics are represented and expressed in a specific source, she would like to receive more news items from the same source. For the news domain it is a challenge not to recommend the same story from different news sources. This approach may also be a solution for this challenge. On one hand, reading several articles from the same source may also give the user a coherent and consistent view of stories. On the other hand, discovering new sources and reading the stories from different sources may give the user a complete, balanced and wider perspective.

The news source ontology that we built is a small scale ontology built for testing the idea of considering news sources in news recommendation process. It is possible to observe more correct relations in a bigger ontology which we consider as a future work. Also any improvements of the quality of the ontology will help to get better results.

The importance of news recommendation which considers the differences and similarities between

news sources looks promising to improve the recommendation quality. On the other hand recommending the news sources itself is something that the news recommenders should consider. This approach that we used to evaluate the similarities between news sources can also be used to find the correlations between news categories and make recommendations considering the news categories where the categorization of news articles is a challenge by itself.

REFERENCES

Cantador, I. and Castells, P. (2009). Semantic contextualisation in a news recommender system. In *Workshop on Context-Aware Recommender Systems (CARS 2009)*.

Capelle, M., Frasincar, F., Moerland, M., and Hogenboom, F. (2012). Semantics-based news recommendation. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, page 27. ACM.

Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F., and Kaymak, U. (2011). News personalization using the cf-idf semantic recommender. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, page 10. ACM.

Gulla, J. A., Ingvaldsen, J. E., Fidjestl, A. D., Nilsen, J. E., Haugen, K. R., and Su, X. (2013). Learning user profiles in mobile news recommendation. pages 183–194.

IJntema, W., Goossen, F., Frasincar, F., and Hogenboom, F. (2010). Ontology-based news recommendation. In *Proceedings of the 2010 EDBT/ICDT Workshops*, page 16. ACM.

Lašek, I. and Vojtáš, P. (2011). Semantic information filtering-beyond collaborative filtering. In *4th International Semantic Search Workshop*.

Lemdani, R., Bennacer, N., Polaillon, G., and Bourda, Y. (2010). A collaborative and semantic-based approach for recommender systems. In *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, pages 469–476. IEEE.

Lops, P., De Gemmis, M., and Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer.

Media, K. (2012). Measuring news consumption and attitudes.

Mobasher, B., Dai, H., Luo, T., and Nakagawa, M. (2001). Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd international workshop on Web information and data management*, pages 9–15. ACM.

Ofcom (2013). News consumption in the uk - 2013 report.

Ozgobek, O., Gulla, J. A., and Erdur, R. C. (2014). A survey on challenges and methods in news recommendation. In *In Proceedings of the 10th International Conference on Web Information System and Technologies (WEBIST 2014)*.

Ozgöbek, O., Shabib, N., and Gulla, J. A. Data sets and news recommendation.

Pang-Ning, T., Steinbach, M., Kumar, V., et al. (2006). Introduction to data mining. In *Library of Congress*.

Peis, E., del Castillo, J. M., and Delgado-López, J. (2008). Semantic recommender systems. analysis of the state of the topic. *Hipertext. net*, 6:1–5.

Rao, J., Jia, A., Feng, Y., and Zhao, D. (2013). Personalized news recommendation using ontologies harvested from the web. In *Web-age information management*, pages 781–787. Springer.

Sandvig, J. J., Mobasher, B., and Burke, R. (2007). Robustness of collaborative recommendation based on association rule mining. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 105–112. ACM.

Sun, X., Kong, F., and Chen, H. (2005). Using quantitative association rules in collaborative filtering. In *Advances in Web-Age Information Management*, pages 822–827. Springer.

Tavakolifard, M., Gulla, J. A., Almeroth, K. C., Ingvaldesn, J. E., Nygreen, G., and Berg, E. (2013). Tailored news in the palm of your hand: a multi-perspective transparent approach to news recommendation. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 305–308. International World Wide Web Conferences Steering Committee.

Wolfe, S. R. and Zhang, Y. (2010). Interaction and personalization of criteria in recommender systems. In *User Modeling, Adaptation, and Personalization*, pages 183–194. Springer.

Zhang, Y. (2005). Bayesian graphical models for adaptive information filtering - ph.d. dissertation.