News Recommendation based on Semantic Relations between Events

Ryohei Yoko, Takahiro Kawamura, Yuichi Sei, and Yasuyuki Tahara, and Akihiko Ohsuga

Graduate School of Information Systems University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan
{r-yokoh,kawamura,sei,tahara,ohsuga}@ohsuga.is.uec.ac.jp

Abstract. In recent years, "News Curation Services" that recommend news articles on the internet to user have been popular. In this study, we propose a new "News Curation Service" that collects and recommends novel articles by using semantic relationships between events in the news articles that a user feels interest. The semantic relationships between events are represented by Linked Data. In order to recommend the news articles to the user, we create search queries by using the sentence structure features. Finally, we collect the news articles on the internet and recommend the articles to the user.

Keywords: Linked Data, News Recommendation System, Information Retrieval

1 Introduction

Recently, web services such as "paper.li"[1] and "The Tweeted Times"[2] that automatically gather news articles, and recommend to users has been popular. The users can easily get information with topicality and novelty by the web services. Such web services are called "News Curation Services". In this study, we propose a "News Curation Service" that recommends news articles with novelty according to interest of a user. We focus on the semantic relationships between terms of news articles that the user shows interest. The semantic relationships between the terms are represented as Linked Data.

2 Related Work

Khrouf et al. [1] recommended event information to convert meta-information (place, time, tag, genre, etc.) represented by Linked Data. They have built a

[1] http://papper.li
[2] http://tweetedtimes.com

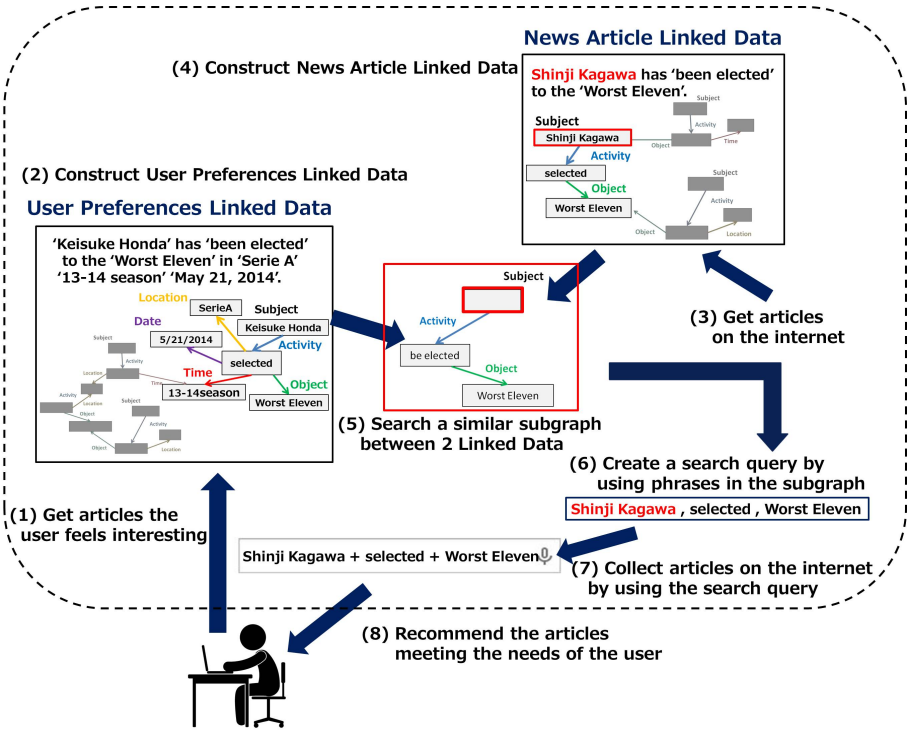

Fig. 1. Summary of Our Approach

hybrid recommendation system by a method using Linked Data and collaborative filtering. In other research, Ostuni et al.[2] has applied user's implicit feedbacks in the service, and recommended top n items by using Linked Data. In recent years, many recommendation systems applying Linked Data have been proposed. However, there is no recommendation system that converts phrases and the semantic relationships within sentences to Linked Data. In addition, use of Bag-of-Words model is a typical method in many recommendation systems, but the model cannot represent the semantic relationships. Our approach can incorporate the interest of the user by the semantic relationships in comparison with the existing methods of using Bag-of-Words vectors.

3 Approach for Creating Linked Data

In this study, we recommend news articles which are collected on the internet according to users' preference. Fig.1 indicates a summary of our method. (1)(2)- first, we collect news articles that a user feels interest and construct "User Preference Linked Data" by each sentence of the article. (3)(4)- next, we gather articles that are candidates of the search queries on the internet to construct

"News Articles Linked Data". Both Linked Data are composed of phrases and semantic relations, that are extracted from sentences of the articles. (5)(6)- finally, if there is a similar subgraph between the two Linked Data, a search query is created as the set of the words of the subgprah.

In order to construct the Linked Data, we extract sentence structures from the articles. We define the sentence structures consists of Subject, Activity, Object, Date, Time, and Location. Each phrase of a sentence is annotated by one of these labels. We represent the sentence structures by using labeled phrases. As an example of Fig. 1, the sentence structure is represented such as "Kagawa→(Activity)→selected" and "selected→(Object)→Worst Eleven" in the Linked Data. We use Conditional Random Fields(CRF) in a similar way as the method of Nguyen et al.[3]. Nguyen et al. annotated the labels in web pages and tweets in Twitter. We recreated a training data set and constructed a learning model for CRF from 100 news articles about soccer written in Japanese in "Sponichi"[3]. "Sponichi" is a famous internet news media in Japan. The average F-measure of labeling with CRF by 10-folds cross-validations is Subject:63.19%, Activity:66.43%, Object:50.06%, Date:91.00%, Time:46.86%, and Location:48.06%.

4 Evaluation

In order to create "News Article Linked Data", we used 14,904 articles from 6/30/2013 to 7/1/2014 in "Sponichi". Likewise, in order to create "User Prefenreces Linked Data", we used 10 articles that a test user feels interest from 152 articles from 8/1/2014 to 8/5/2014 in "Sponichi".

We asked to the test user whether the recommended article is fun(feels fun by the information of the article), novelty(feels new) and serendipity(feels discovery) to each testers. There are eight test users who are our university students. The test users answered in 4 levels: "I thinks so", "I think so a little", "I don't think so a little", and "I don't think so". We recommended top five articles that are searched in Google by the search query created in the previous section. We choose five search queries at random for each test user, so that we recommend 25 articles to a test user. We also defined that a score of the first article is 5 point, the second is 4 point $\cdots$, and the fifth is 1 point. Thus, the perfect score is 15 for each query. In the case that the test users answers "I thinks so", and "I think so little" to the recommended article, the article was regarded as OK, and we added the corresponding point.

We conducted comparison to the method using tf-idf method. We extracted frequent words from 10 articles that the test user feels interest as our proposed method. The search query is created by three of top 10 words calculated by tf-idf method. Fig. 2 indicates the average scores by our method and the tf-idf method. Our method was better than the user tf-idf method. Thus, our approach is effective as a method of news recommendation. In particular, we confirmed that it recommends news articles that are novel to the user.

[3] http://sponichi.co.jp

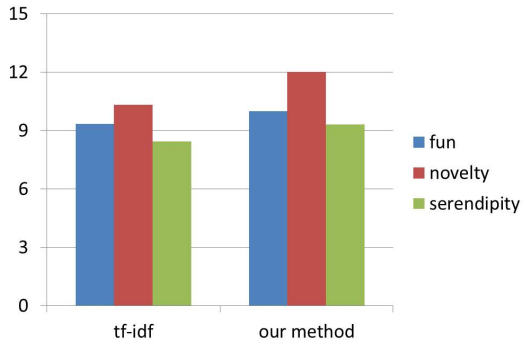

Fig. 2. Comparative Evaluation Results

5 Conclusion and Future Work

In this paper, we proposed a "News Curation Service" by using semantic relationships within sentences. Our approach incorporates interest of a user in comparison with the existing methods. In the future, we will improve accuracy of the labeling.

References

1. H. Khrouf, R. Troncy, Hybrid event recommendation using linked data and user diversity, Proceedings of the 7th ACM conference on Recommender systems, pp.185-192, 2013.
2. Ostuni, Vito Claudio and Di Noia, Tommaso and Di Sciascio, Eugenio and Mirizzi, Roberto: Top-N recommendations from implicit feedback leveraging linked open data, Proceedings of the 7th ACM conference on Recommender systems, pp.85-92, 2013.
3. T.M. Nguyen, T. Kawamura, Y. Tahara, and A. Ohsuga: Self-supervised capturing of users´ activities from weblogs, International Journal of Intelligent Information and Database Systems, Vol.6, No.1, pp.61-76, 2012.
4. Lafferty, John D. and McCallum, Andrew and Pereira, Fernando C. N. : Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning, pp.282-289, 2001.