Towards Serendipitous Research Paper Recommender using Tweets and Diversification

Chifumi Nishioka[1], Jörn Hauk[2], and Ansgar Scherp[3]

[1] Kyoto University Library, Kyoto, Japan
nishioka.chifumi.2c@kyoto-u.ac.jp
[2] Kiel University, Kiel, Germany
[3] University of Essex, Colchester, UK

Abstract. In this paper, we examine whether a user's tweets can help to a generate more serendipitous recommendations. In addition, we investigate whether the use of diversification applied on a list of recommended items further improves serendipity. To this end, we conduct an experiment with $n = 22$ subjects. The result of our experiment shows that the subject's tweets did not improve serendipity, but diversification results in more serendipitous recommendations.

Keywords: Recommender system · Scientific publication · User study.

1 Introduction

Various works have developed recommender systems for research papers to overcome the information overload problem. Most of the previous works have focused on the accuracy of recommendations. However, several works argue that there are important aspects other than accuracy [4]. One of these aspects is *serendipity*, which is concerned with the novelty of recommendations and in how far recommendations positively surprise users [2].

In this paper, we study a research paper recommender system focusing on serendipity. Specifically, this paper conducts an experiment to investigate the influence of user's tweets and diversification to deliver serendipitous recommendations. The experiment is composed of three factors. In the first factor *User Profile Source*, we compare the two sources of a user's own papers vs. the user's tweets. We assume that user's tweets produce recommendations that cannot be generated based on papers, since researchers tweet about very recent developments and interests that are yet not reflected in their papers. In the second factor *Text Mining Method*, we apply three different methods for computing profiles of candidate items (i.e., research papers) and user profiles. In the third factor *Ranking Method*, we compare two ranking methods: classical cosine similarity and the established diversification algorithm IA-Select [1]. IA-Select ranks candidate items with the objective to diversify recommendations in a list. Since it broadens the coverage of topics in a list, we assume that IA-Select delivers more serendipitous recommendations. The result of the experiment reveals that users' tweets did not improve the serendipity, but IA-Select delivers more serendipitous recommendations.

2 Experimental Factors

In this paper, we build a content-based recommender system along with the three factors *User Profile Source*, *Text Mining Method*, and *Ranking Method*. It works as follows: (a) Candidate items of the recommender system (i .e., research papers) are processed by a text mining methods and paper profiles are generated. (b) A user profile is generated based on his/her user profile source by the same text mining method, which is applied to generate paper profiles. (c) One of the ranking methods determines the order of recommended papers. In the following paragraphs, we describe the details of each factor. The three factors described above result in total in $2 \times 3 \times 2 = 12$ strategies.

User Profile Source In this factor, we compare the following two data sources that are used to build a user profile.

- **Own papers:** As baseline, we use the own papers of the users as Sugiyama and Kan [8] did.
- **Twitter:** In contrast to the user's papers, we assume that using tweets provide more serendipitous recommendations, since researchers tweet about their most recent interests.

Text Mining Method For data sources, we apply a profiling method using one of three text mining methods:

- **TF-IDF:** We use TF-IDF since it is often used in recommender systems as baseline [3].
- **CF-IDF:** Concept Frequency Inverse Document Frequency (CF-IDF) [3] is an extension of TF-IDF, which replaces terms with semantic concepts from a knowledge base.
- **HCF-IDF:** Hierarchical Concept Frequency Inverse Document Frequency (HCF-IDF) [6] is an extension of CF-IDF. It applies a propagation function [5] over a hierarchical structure of a knowledge base to give a weight to concepts in higher levels. Thus, it identifies concepts that are not mentioned in a text but highly relevant.

Ranking Method Finally, we rank all candidate items to determine which items are recommended to a user. In this factor, we compare two ranking methods: cosine similarity and diversification with IA-Select [1].

- **Cosine similarity:** As baseline, we employ a cosine similarity. Top-k items with largest cosine similarities are recommended.
- **IA-Select:** We employ IA-Select [1] for serendipitous recommendations. IA-Select diversifies recommendations in a list to avoid suggesting similar items together. The basic idea of IA-Select is that it lowers iteratively the weights of features in the user profile, which are already covered by papers already selected for recommendation.

3 Evaluation

Procedure Along with the previous work [7], we have implemented a web application where human subjects evaluate the twelve recommendation strategies described above. First, subjects input their Twitter handle and their name. Based on their name, we obtain the content of their papers by mapping them to the ACM-Citation-Network V8 dataset (see below). The top-5 recommendations are computed for each strategy. Thus, each subject evaluates $5 \cdot 12 = 60$ items as "interesting" or "not interesting". Subjects can directly access and read the research paper by clicking the link.

Datasets As research papers, we use the ACM citation network V8 dataset[4]. From the dataset, we use 1,669,237 papers with title, author, year, venue, and abstract. As a knowledge base for CF-IDF and HCF-IDF, we use the ACM Computing Classification System (CCS) [5].

Subjects Overall $n = 22$ subjects were recruited. The subjects published on average 1256.97 tweets (SD: 1155.8). Regarding research papers for user profiling, on average a subject has 11.41 own papers (SD: 13.53).

Metric To evaluate the serendipity of recommendations, we use the Serendipity Score (SRDP) [2]. It takes into account both unexpectedness and usefulness of candidate items, which is defined as: $SRDP = \sum_{d \in UE} \frac{rate(d)}{|UE|}$. UE denotes a set of unexpected items that are recommended to a user. An item is considered as unexpected, if it is not included in a recommendation list computed by the primitive strategy. We use the strategy Own Papers × TF-IDF × Cosine Similarity as a primitive strategy. The function $rate(d)$ returns an evaluation rate of an item d given by a subject. If a subject evaluates an item as "interesting", it returns 1. Otherwise, it returns 0.

4 Result and Discussion

Table 1 shows the results of the twelve strategies in terms of SRDP. Since we use the strategy Own Papers × TF-IDF × Cosine Similarity as a primitive strategy, mean is .00 for the strategy. An ANOVA is conducted to detect significant differences between the strategies. The significance level is set to $\alpha = .05$. Applying a Muchly's test detects a violation of sphericity ($\chi^2(54) = 80.912$, $p = .01$). Thus, a Greenhouse-Geisser correction with $\epsilon = 0.58$ is applied. The ANOVA reveals significant differences between the strategies ($F(5.85, 122.75) = 3.51$, $p = .00$). Shaffer's modified sequentially rejective Bonferroni procedure reveal significant differences between the primitive strategy and one of the other strategies.

The results of our experiment showed that tweets do not improve the serendipity. As shown at the rightmost column in Table 1, tweets deliver unexpected

[4] https://lfs.aminer.org/lab-datasets/citation/citation-acm-v8.txt.tgz
[5] https://www.acm.org/publications/class-2012

recommendations to users. However, only a small fraction of these serendipitous recommendations were interesting to the users. The results show further that the IA-Select algorithm produces serendipitous recommendations. Thus, IA-Select can be used in a research paper recommender to improve serendipity.

Table 1. SRDP and the number of unexpected items of the twelve strategies.

	Strategy			**SRDP**	**|UE|**
	Text Mining Method	**Profiling Source**	**Ranking Method**	**M (SD)**	**M (SD)**
1.	TF-IDF	Own Papers	IA-Select	.45 (.38)	2.95 (1.05)
2.	CF-IDF	Twitter	CosSim	.39 (.31)	4.91 (0.29)
3.	TF-IDF	Twitter	IA-Select	.36 (.29)	4.91 (0.43)
4.	CF-IDF	Twitter	IA-Select	.31 (.22)	4.95 (0.21)
5.	CF-IDF	Own Papers	CosSim	.26 (.28)	4.91 (0.29)
6.	CF-IDF	Own Papers	IA-Select	.25 (.28)	4.91 (0.29)
7.	HCF-IDF	Own Papers	IA-Select	.24 (.22)	4.95 (0.21)
8.	HCF-IDF	Twitter	CosSim	.22 (.28)	5.00 (0.00)
9.	TF-IDF	Twitter	CosSim	.20 (.24)	4.95 (0.21)
10.	HCF-IDF	Twitter	IA-Select	.18 (.21)	5.00 (0.00)
11.	HCF-IDF	Own Papers	CosSim	.16 (.18)	5.00 (0.00)
12.	TF-IDF	Own Papers	CosSim	.00 (.00)	0.00 (0.00)

5 Conclusion

We have investigated whether tweets and IA-Select deliver more serendipitous recommendations. Our online experiment reveals that tweets do not improve the serendipity of recommendations, but IA-Select does. This insight contributes to the development of future recommender systems in such a sense that a provider can make more informed design choices for the systems and services developed.

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying search results. In: WSDM. pp. 5–14. ACM (2009)
2. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In: RecSys. pp. 257–260. ACM (2010)
3. Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F., Kaymak, U.: News personalization using the CF-IDF semantic recommender. In: WIMS. ACM (2011)
4. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. TOIS **22**(1), 5–53 (2004)
5. Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A.: User interests identification on Twitter using a hierarchical knowledge base. In: ESWC. Springer (2014)
6. Nishioka, C., Große-Bölting, G., Scherp, A.: Influence of time on user profiling and recommending researchers in social media. In: i-KNOW. ACM (2015)
7. Nishioka, C., Scherp, A.: Profiling vs. time vs. content: What does matter for top-k publication recommendation based on Twitter profiles? In: JCDL. pp. 171–180. ACM (2016)
8. Sugiyama, K., Kan, M.Y.: Scholarly paper recommendation via user's recent research interests. In: JCDL. ACM (2010)