SMART Journalism: Personalizing, Summarizing, and Recommending Financial Economic News

Maya Sappelli, Dung Manh Chu, Bahadir Cambel, David Graus, Philippe Bressers

FD Mediagroep
Prins Bernhardplein 173, 1097 BL
Amsterdam, The Netherlands

Abstract

FD Mediagroep (FDMG[1]) is the leading information provider in the financial economic domain in the Netherlands. FDMG operates "Het Financieele Dagblad" (FD) a daily financial newspaper, similar to the Financial Times. In addition, FDMG operates the daily all-news radio station "Business News Radio" (BNR). As we have a wide variety of users with various backgrounds and interests, we believe that digital media (both news articles and radio) should be personalized to match the interests of a particular customer. We are therefore working on personalization of FDMG's digital media:

- Personalized news: Recommendations and personalized summaries of news articles that match the reading preferences and interests of our readers

- Personalized radio: A non-linear radio experience with radio snippets that match the listener's interests

In both personalized news and personalized radio we are looking not only at introducing recommender systems but also at personalized ways to present the information using automated summarization (news) and audio segmentation (radio) methods

In the digital era, the amount of available information and the pace at which it comes available to users increase rapidly. It is difficult to keep track of all information, while at the same time many of us want to keep up to date on specific topics that interest us.

In the news domain, landing pages and radio shows are put together by editors. While much effort is spent in better understanding the preferences and interests of our readers and listeners, a single landing page that aims to serve all users, cannot take into account the variety of specific backgrounds and subtle deviations in interests across users. We believe that for digital media (both news articles and radio) it is important to both personalize the selection of content items to better match the interests of the users, and to take the users' contexts into account. For example, the duration of radio fragments and the length of article summaries should be reduced when little time is available, whereas users with more available time should receive longer summaries and fragments. For this purpose we are researching and developing personal news and radio applications, as a first step towards personal and context-aware digital media delivery.

Personalized News

FDMG's daily financial newspaper "Het Financieele Dagblad" (FD) publishes on average 70 articles a day, while only about a third of these articles makes it to the main landing page, typically remaining visible for only a limited amount of time. This means that many articles on these pages are not easily found. Recommender systems present opportunities for serving these less accessible articles to users through e.g., personalized landing pages, moreover they can make 'old' news relevant again.

Through the SMART Journalism project we aim to create a personalized landing page for each user, using a hybrid content-based and collaborative-filtering recommendation approach. Collaborative filtering approaches have proven very effective (Ricci et al., 2010), but suffer from the cold-start problem. This is particularly relevant for the news domain, as it is important to include fresh news fast in personal landing pages (Liang, 2017). To alleviate the cold start problem, content analysis will help to recommend fresh articles. We aim to analyze the topic of articles to match them to users' profiles. In addition to this content-based recommendation, other aspects that we aim to study include recency, popularity, editorial relevance, diversity (Zhou et al., 2010), and serendipity (Bordino, Mejova, and Lalmas, 2013) of our personalized pages. As a FDMG-specific feature, we are additionally looking into group-based recommendations. At "Het Financieele Dagblad," we have two types of user subscriptions: (i) individual subscribers, and (ii) organization-based subscribers where a subscription grants a large number of an organization's employees access to our news site and app. We have seen in the past that employees within an organization typically share similar reading interests, particularly with news relevant to the organization's sector. Grouping users by their organization and generating organization-specific recommendations is are promising ways of incorporating additional contextual data into recommendations (Masthoff, 2011).

As a proof of concept of the benefit of recommendation algorithms in the news domain, we have run our first experiments with a personalized weekly newsletter. As opposed

[1]http://fdmediagroup.com/

to using the most popular articles, we compose personalized newsletters using collaborative filtering-based recommendation, based on users' past reading behavior. More specifically, we collect reading data of the last four weeks to construct an item-to-item similarity matrix. Using personalized newsletters, we now send out on average 180 articles each week, compared to an average of 10 articles using the previous popularity-based method. Among the 180 articles sent per week, on average one third of the articles is clicked and read. This shows that we are able to serve a larger portion of our articles to our readers. Moreover, clickthrough rates are significantly higher for the personalized newsletters (29% increase) as compared to the previous popularity-based newsletters, suggesting that the users are more satisfied with the suggestions given to them.

Personalized Summaries

In addition to the recommendation of news articles to users, we are exploring how articles can best be presented. Here the context of a user plays an important role. If a user is on her way to work, she has different information needs than when she is at home on the weekends. By using automated summarization methods we aim to create summaries that match a user's reading preferences and context. We aim to, e.g., present shorter summaries for readers who have little time, infographics for readers that do not want to read too much, more in-depth and detailed summaries for readers that are knowledgeable on the article's topic, and summaries with additional explanations for readers who are relatively new to the topic.

Automated summarization has seen some promising successes recently, largely through the rise of deep learning in the natural language processing domain (Paulus, Xiong, and Socher, 2017; Nallapati, Xiang, and Zhou, 2016). However, the task of *personalizing* these summaries is challenging. And even though there are some successes there as well (Díaz and Gervás, 2007), they have not yet profited from the recent improvements in summarization. In addition to query-based and user-based personalization, we will be looking at aspects such as tone of voice and content style. Furthermore, adapting the methods that are primarily developed and evaluated in English to the Dutch language may bring additional unforeseen challenges, particularly because there are no relevant Dutch summarization datasets.

Personalized Radio

"Business News Radio" (BNR)[2] is FDMG's all-news radio station focusing on enterprising, ambitious people. At BNR we produce daily radio news and podcasts in the business and financial economic domain. Currently we provide a simple tool to search through a limited selection of our archive of broadcasted news.

However, we aim for more effective and innovative methods of bringing our rich archive of audio content to our listeners. More specifically, as searching through an archive of full-length recordings (typically shows have a duration of 1 to 2 hours) for specific topics may be a bit cumbersome,

and as sometimes only a smaller fragment or snippet from a show is interesting to a user, we are developing methods for automated segmentation of radio shows, based on the topics that are discussed, in order to serve the users more relevant content. Instead of recommending a full-length recording of a show, we aim to recommend a set of relevant fragments and snippets from different shows, to create personalized and topic-based playlists. To make this possible, we apply amongst other things jingle detection (Pinquier and André-Obrecht, 2004) and topic detection (Wang and Manning, 2012).

Discussion

Given our plans for personalizing news and radio, it is also important to reflect on potential risks. One such risk is the possibility that users might end up in a "filter bubble" through very precise inference of a user's interests. However, experimental results from our personalized newsletter suggests that recommendation technology can in fact diversify the set of recommended articles, compared to recommendations by editors or popularity-based recommendations. This is consistent with the findings by Nguyen et al. (2014) and Zuiderveen Borgesius et al. (2016).

In addition, recommendation and personalization require the collection of user data (i.e., reading behavior) to estimate users' preferences and interests. At FDMG we strive to protect the privacy of our users. This means that the user should always be in control over the data we collect. A likely risk is that this control negatively impacts the effectiveness of the personalization. Group personalization rather than individual personalization could be one method to alleviate privacy concerns while maintaining the possibility to personalize.

Acknowledgements

References

Bordino, I.; Mejova, Y.; and Lalmas, M. 2013. Penguins in sweaters, or serendipitous entity search on user-generated content. CIKM '13, 109–118. New York, NY, USA: ACM.

Díaz, A., and Gervás, P. 2007. User-model based personalized summarization. *Information Processing & Management* 43(6):1715–1734.

Liang, Y. 2017. An ensemble approach for news recommendation based on contextual bandit algorithms. Master's thesis, TU Delft.

Masthoff, J. 2011. Group recommender systems: Combining individual models. In *Recommender systems handbook*. Springer. 677–702.

Nallapati, R.; Xiang, B.; and Zhou, B. 2016. Sequence-to-sequence rnns for text summarization. *CoRR*.

---

[2]http://www.bnr.nl

[3]https://newsinitiative.withgoogle.com/dnifund/

Nguyen, T. T.; Hui, P.-M.; Harper, F. M.; Terveen, L.; and Konstan, J. A. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. WWW '14, 677–686. ACM.

Paulus, R.; Xiong, C.; and Socher, R. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

Pinquier, J., and André-Obrecht, R. 2004. Jingle detection and identification in audio documents. In *Acoustics, Speech, and Signal Processing (ICASSP'04)*, volume 4, iv–iv. IEEE.

Ricci, F.; Lior, R.; Shapira, B.; and Kantor, P., eds. 2010. *Recommender systems handbook*. Springer US.

Wang, S., and Manning, C. D. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. ACL '12, 90–94.

Zhou, T.; Kuscsik, Z.; Liu, J.-G.; Medo, M.; Wakeling, J. R.; and Zhang, Y.-C. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *PNAS* 107(10):4511–4515.

Zuiderveen Borgesius, F.; Trilling, D.; Moeller, J.; Bodó, B.; de Vreese, C. H.; and Helberger, N. 2016. Should we worry about filter bubbles? *Internet Policy Review* 5(1).