Interpreting News Recommendation Models

Blaž Fortuna
Ghent University – iMinds, Belgium
Jožef Stefan Institute, Slovenia
blaz.fortuna@ijs.si

Pat Moore
Bloomberg L.P., USA
pmoore26@bloomberg.net

Marko Grobelnik
Jožef Stefan Institute, Slovenia
marko.grobelnik@ijs.si

ABSTRACT
This paper presents an approach for recommending news articles on a large news portal. Focus is given to interpretability of the developed models, analysis of their performance, and deriving understanding of short and long-term user behavior on a news portal.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: On-line Information Services – *Web-based services*

General Terms
Algorithms, Human Factors.

Keywords
News recommendation, Personalization, Learning to rank.

1. INTRODUCTION
Widespread use of the web brought changes to how news is being consumed. In the past, newspapers or magazines would contain articles covering a particular topical and/or geographical area. The articles would be manually selected and arranged on the paper by editors to cover the important news since the last issue, provided the limited real estate of the paper. On the web, news portals can provide access to recent events in near real-time, resulting in multitude of articles available for each particular event. News portals also provide richer navigation mechanisms compared to traditional newspapers, allowing their readers to focus on their topics of interests, while disregarding others.

In this paper we analyze web server access logs of a large online news publisher to identify readership patterns on the web. In particular, the analysis is done by first developing a model, which can be used to predict most likely articles to be read by a particular user, followed by analyzing what are the most important features and interpreting the learned model.

The techniques and approaches presented in this paper build largely on similar work done in the area of news recommenders [1]. This paper focuses on the interpretation of the developed models, and the contribution of various observable modalities on their performance.

The paper is organized as follows. First we present the dataset and preprocessing methodology. This is followed by description of features used in the model and the paper concludes with the analysis of the trained models.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.

http://dx.doi.org/10.1145/2740908.2742006

2. DATASET
The dataset used in this paper consists of web server access logs obtained from a large online news portal for the period of one month. The portal publishes daily around a thousand news articles, and has one million page views per day from 300,000 unique visitors. Experiments in this paper only use the access logs for article pages, disregarding pages such as homepage and section fronts.

All the article pages, which occur at least once in the access logs, were crawled and article title, content, and publish date were extracted.

3. MODEL AND FEATURES
We use two layers to assemble the user models. In the first layer, each article is represented using one of several feature spaces, which can be roughly assigned to two groups: content and collaborative. In the second layer, each user $u \in U$ is represented as a set of all the articles the user read.

Content modalities for representing articles correspond to features, which can be directly or indirectly extracted from the content of the article. This includes the words from the article title and body (referred to as *Content*), and the category (referred to as *Categories*) of article. We use bag-of-words [2] representation for representing title and body. For categories, analogues "bag-of-categories" is used.

Collaborative modalities for representing articles correspond to features, which can be extracted just by observing visit logs, and disregarding the article's content. The most basic collaborative feature used is the popularity of an article (number of visits). This would correspond to the "most popular" lists frequently seen on news portal, and will be referred to as *Popularity*.

There are also two more complex sets of collaborative feature. First, for each article, a weighted list of co-visited articles is derived. Articles a_1 and a_2 are counted as co-visited each time they appear together within a user session. Here we use "bag-of-co-visited-articles" representation and refer to it as *Co-visits*. Second, for each article we keep a list of all the users, which read the article. Here we use "bag-of-users" representation and refer to it as *Users*.

The following procedure is used, to predict the article most likely to be read next by a user u at time t_0. First, a set of potential articles A is assembled, by selecting all the articles between $t_0 - t_{wnd}$ and t_0, minus the articles already read by the user.

Second, each article $a \in A$ is assigned a score for each of the feature modalities. In the case of *Popularity*, the score $f_P(a)$ corresponds to the number of visits. For all other modalities, the score $f(a, u)$ is computed as an average of cosine similarities between the article a and the articles from the users' profile $(a \in u)$. The end result of the second step is a feature vector

$x_a = (x_1, \ldots, x_n)$ for each article, with features covering different aspects of similarity between the article and the user.

As a special case, scores $f(a, u)$ can also rely only on the last read article by the user. Such features will be referred to as *recent features* in the experimental section, and are used alongside normal features (concatenated to the feature vector x_a).

Finally, each article from the time window is assigned a score, by computing a weighted combination of its features $w^T x$. The weights w are computed using RankSVM [3] over the visit logs data from the first half of dataset, and high score corresponds to higher likelihood of article being read by the user in the near future.

Please note that the features and the approach were selected as to allow for real-time updates and recommendation. This would be harder were we to use more sophisticated collaborative methods, i.e. techniques based on the user-item matrix decomposition.

4. EXPERIMENTS

The experiments have two major parts. In the first part we try to estimate what is the right time window t_{wnd} based on average age of articles when read by the users. In the second part we analyze the predictive performance of each single feature and their combinations.

Time window selection can have a significant impact on the accuracy of predictions, as outlined in Section 3. We also expect the time window to vary largely between and within news portals, depending on the news domain. For example, financial news would have much shorter shelve-life compared to opinions or larger overview articles. The median age of consumed articles as observed in access logs, used in the following experiments, was between 6 to 8 hours during the week. It increases to 12 hours on Saturday and 24 hours on Sunday. On average, double the median (e.g. 15 hours on weekday) would cover the age of roughly 90% of all the articles read, and as such was used as the time window.

In the experiments, the users were split into four groups, based on the number of articles in their profile. There were also two sets of feature vectors: with and without recent features.

The experiments were done as follows. First, a random timestamp t_0 is selected from the second half of November. Second, a user is selected, which requested an article 10 minutes before and after t_0. Thirdly, a ranked list of predicted articles is assembled, using the procedure specified in Section 3. Finally, if the article read right after t_0 is among top four articles from the assembled list, the prediction is scored 1, and 0 otherwise. The presented results are the average of this score over 100,000 tests for each group.

In the first experiment, each modality was tested individually, by ranking the articles according to the modality's corresponding feature. For example, in the case of *Content,* this results in ranking articles by cosine similarity with the user's profile. The results in Table 1 show high baseline set by *Popularity*, which is partially due to high prominence of "Most popular" list on the news portal. It can be seen that the performance of *Content* and *Co-visit* features does not increase with the number of articles in the user profile.

In the second experiment, RankSVM model was used to learn weights for combining feature sets. We trained a separate model for each user group. The performance of several feature set combinations is shown at the bottom half of Table 1. First, the difference between the feature sets is the inclusion of *Users* feature, which is computationally the most expensive. Second, the difference is the inclusion of recent features. Both feature sets were found to provide considerable boost to the performance.

We can check the importance of each feature set by checking corresponding weights assigned by RankSVM. The weights are shown in Table 2. First, it can be seen that the *Popularity* influence drops as the user profile grows, and becomes negative for users with more than 10 articles in their profile. *Users* feature is the most informative, resulting in high weights across all groups. Both *Content* and *Categories* are positive when averaged across whole user profile, but become negative when used only on the last read article. This shows that users are not really interested in more articles within the same narrow topic in a single session (e.g. articles about price of Gold), but maintain focused on their topics when average over longer time period.

Table 1. Performance of single modalities (top) and of combined feature sets (bottom).

#articles	1	2-10	11-50	51-
Popularity	0.13	0.15	0.19	0.19
Content	0.06	0.07	0.09	0.07
Categories	0.06	0.10	0.16	0.21
Co-visits	0.14	0.16	0.14	0.12
Users	0.19	0.23	0.27	0.30
SVM [no users]	0.19	0.19	0.22	0.22
SVM [no users, recent]	0.19	0.21	0.29	0.32
SVM [with users]	0.21	0.23	0.27	0.31
SVM [with users, recent]	0.21	0.24	0.32	0.37

Table 2. Weights for combining feature sets

#articles	1	2-10	11-50	51-
Popularity	1.28	0.58	**-0.10**	**-0.30**
Content	0.17	0.85	1.09	0.66
Categories	0.31	0.58	0.72	1.09
Co-visits	0.67	0.39	**-0.40**	**-0.29**
Users	4.45	10.33	11.59	8.56
Content [recent]	0.17	**-0.06**	**-0.47**	**-0.30**
Categories [recent]	0.31	**-0.04**	**-0.20**	**-0.19**
Co-visits [recent]	0.67	1.03	1.35	1.23
Users [recent]	4.45	1.95	0.24	1.05

5. CONCLUSIONS

In this paper we presented an approach for modeling users for news recommendation scenario. We evaluated the approach on a large dataset, comprising one month of access logs from a large news portal. The results show the importance of collaborative features. Further, it can be seen that in long term the users stay focused in their topics of interest, but prefer diversity within one session.

6. ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the FP7 project XLike (ICT-2011.4.2 288342).

7. REFERENCES

[1] Lei Li; Ding-Ding Wang; Shun-Zhi Zhu; Tao Li. *Personalized news recommendation: A review and an experimental investigation.* Journal of computer science and technology. Vol 26: No: 5, pp.754-766, 2011.

[2] Manning, C.D.; Schutze, H. *Foundations of statistical Natural Language Processing* (MIT Press, 1999).

[3] L. Tie-Yan. *Learning to Rank for Information Retrieval.* Foundations and Trends in Information Retrieval. Vol. 3: No 3, pp. 225–331, 2009.