

# News recommendation using implicit feedback

Vladimír DEMČÁK\*

*Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
vlado.demcak@gmail.com*

**Abstract.** This work deals with news recommendation focusing on the using implicit feedback between user and system. Our solution is focused mainly on content based recommendation technique and English language. We have designed special part of graphical user interface named as "brick" which represents one news in our news portal. By default brick is showing title, image and short description of article. Text description is blurred and it become readable after user put mouse cursor over description. Using this feature we try to use mouse cursor as an equivalent to the board pointer for finding out where and what user reads. We want to determine active of reading articles through special reading mode which we have designed. Our goal is also to take advantages of social networks where users share and express their interests and combine this technique in recommendation. Social networks surely provide very important statistics when we know number of shares, comments or likes for specific news. Our solution is web application that consists of client and server part.

## 1 Introduction

News is term that can have many different meanings. In this work we consider term "News" as is defined in [7]: *"short and brief textual information that has to have public meaning and must inform about some stories or news contexts.* News portals produce every day more and more articles in different topics, areas and with various qualities. News portals try to create unique content for every visitor through recommendations. Process of recommendation is still very problematic area that needs to bring something new and innovative especially in domain of news.

Automatic creation of personalized content is a difficult process. It demands to solve some complex problems like few or none information about user or items (known as *cold-start problem*), selection of recommendation technique for specific domain, model creation for user, natural language processing and export structural information which is understandable for computers. Cold-start problem is one of the most common problem in recommendation. It can be noticed while processing new items or new users [6]. We can eliminate this problem with combination of different recommendation techniques (hybrid approach in recommendation). Or sometimes is possible to take user's previous activities from other systems that he has been using in the past or still uses (e.g. social networks

---

\* Master study programme in field: Information Systems  
Supervisor: Professor Pavol Návrat, Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

like Facebook or Twitter). Recommendation is still really problematic area and new knowledges and approaches are analysed and discussed on conferences like RecSys<sup>1</sup> or RSSEs<sup>2</sup>.

## 2 Related Work

News articles are in natural language and represent unstructured information. These informations carry important informations that are useful in generating of recommendation. One of the examples is named entities. Hence before saving news we need to analyse meta-data and properly organize preprocessed news [4]. Preprocessing includes actions like classification and text pre-processing. Pre-processing is process when system processes raw data into enough interpreting format [5]. In context of natural language pre-processing we are talking about few steps as tokenization, text processing or text classification [2]. Tokenization divides words in raw text. Then comes removing stop words - words like prepositions or any useless words and after that follows e.g. counting TF-IDF weights for documents. One of goals in text processing is classification - to assign news to specific category. Text classification often uses classification model which is known as *classifier*. Process of classification consists of two main steps [2]. First one is building model based on train dataset. Dataset is usually a big set of news that are manually assigned into specific topics by human expert. The second step is applying previously created model on real data. The most popular methods are SVM (Support Vector Machine) or Naive Bayes.

In recommendation domain there are many standard techniques. The most used are collaborative filtering and content based recommendation. Collaborative filtering is aimed at group of things or users. The main idea is to find out similarity between entities and try to recommend things that are interesting for similar users and have not been in contact with user for those system is generating recommendations. This technique is also divided into two approaches. Memory-based collaborative filtering (e.g. what user read in the past) and model-based collaborative filtering. Models are usually created with algorithms like Bayes networks, clustering or rule-based systems [3]. Collaborative filtering is good technique when the number of users and items is stable [4]. However domain of news is quickly changing and the most important things in news is novelty. So the better approach seems to be content based recommendation in comparing to the collaborative filtering.

Day in week (S/L)	Error clicks (S/L)
Part of a day (morning/ day / night) (S/L)	Size of news (W)
Key words (S/L)	Reading time(W)
News novelty (S/L)	Read again(W)

Table 1. Attributes considered in creation of SWL model in research [4]

Implicit feedback for better analysing user's interests is used in research [4]. Authors presented model SWL (Select-Watch-Leave) which is the main part of generating recommendation. Every characters represents the part of model that influences (S/L - selection or leaving news, W - watching/reading article). The whole process is illustrated in their work and also described in few steps. The significant step is logging user activities and actualization of user profile (his interests) based on his implicit feedback. In table 1 we can see which attributes and information they were analysing. In brackets are characters represent which part of model it influence. Research [1] describes interesting system architecture for generating recommendation based on hybrid technique that involves collaborative filtering and content-based recommendation. They also considered the most popular news

<sup>1</sup> RecSys - ACM Recommender Systems - online: <http://recsys.acm.org>

<sup>2</sup> Recommendation Systems for Software Engineering (RSSEs) online: <https://sites.google.com/site/rssersearch/>

in global view, random news but mostly monitoring user's clicks. Topic modelling is implemented with Latent Dirichlet Allocation (LDA) method in background.

### 3 Our solutions and method

As the title says we are focused on implicit feedback. It is without doubt very useful indicator what is user interested in. As in many researches we analyse number of clicks or size of news. Below we introduce some major parts, method and approach in our solution.

#### News brick

We enhance existing approaches with new element that we call "News brick" in the topic sections. Using this feature we try to use mouse cursor as a equivalent to the board pointer for finding out where and what user reads. Brick represents one news. We assume that when user visit news portal he just take his sight on specific news and we have to use eye tracker or something like this to find out where he is looking at. We want to force user to perform small and easy action (moving with mouse) to resolve this problem. In default state brick shows title, image and short description of article. Description is blurred and become readable when user put mouse cursor over description.



Figure 1. Brick with blur effect. Default state on the left side and mouse cursor over text on the right side

#### Obtaining news

We are trying to create recommendation system that works with real and actual news. System has to download and analyse actual news. This activity is repeatable and automatic on server side. We considered and implemented two ways how to download and access to news on the news portals.

- *Web-crawler* - program which collects web pages on the Internet. In our case pages with news.
- *RSS feeds* - provided by news portals as a source of published news.

Web crawler starts on some web page (defined in database) and then visit every web http links that meet the conditions (not visit images, .js etc.). RSS feeds are defined in database. News are classified (assigned to topics) with system AlchemyAPI which is a part of the project IBM Watson<sup>3</sup>.

<sup>3</sup> About US — AlchemyAPI, online: <http://www.alchemyapi.com/about-us>

### Reading mode

System provides special reading mode on graphical user interface. We also want to determine active of reading news articles through special reading mode. This mode is basically inserting blocks (e.g. images or advertisements) in order to divide text into parts. User is able to read only part of the news (visible only few lines of text) and user is forced to scroll down that means he is at screen for that moment. We assume that this action is not very hard to perform and active reading is very meaningful. Figure 2 shows comparison of reading mode and raw text of article.



Figure 2. Comparing "Reading Mode" on left and raw text of article on right.

### News content and global view on social network

As content we consider mainly named entities like people, organisations and locations. This is for us the base information what is article about. With these informations we find similar articles based on content. Except content-based recommendation we consider also **global view** about articles. We are focused on statistics from social network Facebook. Our system uses Graph API which is simple interface for obtaining data for specific user and news from Facebook. These statistics are obtained from Facebook automatically and are displayed in small chart in reading mode. We consider statistics like number of shares, number of comments and number of likes for specific news.

Every user is identified by unique automatically generated cookie which is stored on client machine and on backend side. Unique identifier is created as combination of actual IP address and randomly generated string. User models is build as scoring vector for entities and sections. These vectors are not of fixed size and size can be changed in time. We also consider "strong" and meaning of different activities. For example reading brick is less important then click on brick in case when we talking about indication of interests.

Vector represents objects with structure:  $v = \{(entity_x, num.of\ news(a), weight(b), \dots)\}$ , where  $b$  is consideration of "strong" for activities that we described few lines above. Then we need to select entity for recommendation but not the same entities for every recommendation. So we use probability of selection based on weights in user model vector as:  $P_x = \frac{b_x}{b_{max}} P_{max}$  ( $P_{max}$  is maximal probability set for entity with the biggest value of ( $b$ )).

## 4 Experiments

We have implemented recommendation prototype and made some experiments. In this part of article we will discuss some of knowledges and informations that we identified. We compared two proposals of obtaining news from news portals like BBC - Web crawler and RSS feeds. Using RSS feed as source for news seems like the better solution. There are two main reasons why it is better. The first one is that we can track statistics of news on social networks from the beginning. Second advantage is that we know that http reference in RSS feeds are linkage to textual news and we do not have to

identify if page is really article. On the 3 we can see comparison of two discussed approaches for tracking information about sharing or commenting on social network Facebook used for global view in recommendation.

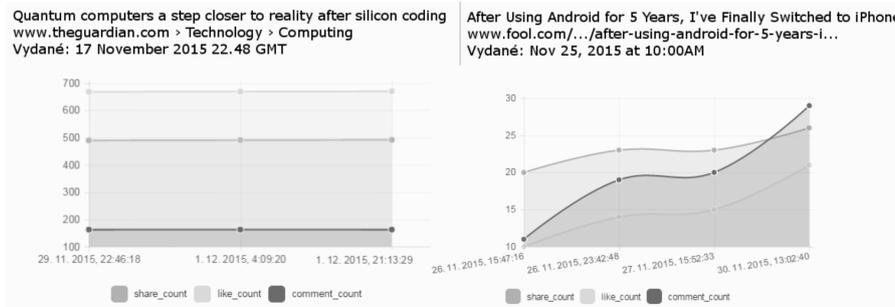


Figure 3. Comparison of tracking news obtained by Web crawler and with RSS Feeds

In first experiment we were working with about 4000 articles collected during 4 days from different RSS feeds (mainly from BBC). As default state system displayed following sections: law govt and politics, technology and computing, art and entertainment, sports, shopping, science and also TOP news based on statistics of global view (data from social network) in last 3 days. We found out that section TOP news should be displayed as a default section because our participant did not notice that. Participant have been using system for about 30 minutes. First action what he did was click on technology and computing section and he started reading. Then put mouse cursor over description in bricks, click on brick and system displayed reading mode of article. System recorded following actions:

- 1, *readingBrick - Uber trial coloured lights to make cars stand out in traffic (11341ms)*
- 2, *readingBrick - Fitbit is the most popular wearable of 2015 as sector grows almost 200pc (14470ms)*

- 3, *clickBrick - Fitbit is the most popular wearable of 2015 as sector grows almost 200pc.*

Then the system displayed following sections (in second session): technology and computing, consumer electronics, beauty, mobile phones, men's health, space technology.

Second experiments was executed on 7 people and we wanted to find out if users put mouse cursor over bricks (blur effect was turned off). Only 1 participant put mouse over brick and the others did not move with mouse.

## 5 Conclusions and future work

In our solution we tried to use mouse cursor to find out where user watches. We presented term "News brick" that includes title, image and short description of article. Text of this entity is blurred by default and user is forced to put mouse over brick. In experiment we identified that about 75% of our respondents did not moved with mouse while they were reading sections on news portal. We also implemented two ways of obtaining news from news portals. We compared web crawler and RSS feeds. Experiment showed as better solution to use RSS feed and then parse web page that contains specific article.

In near future we will focus our work on offline and online evaluation and next experiments. Offline evaluation will be used as phase of settings. With this action we want to find out the best values for specific actions (for implicit feedback). After this phase we will execute online evaluation and we will execute survey about interactions with portal and generated recommendations.

## References

- [1] Garcin, F., Faltings, B.: PEN Recsys: A Personalized News Recommender Systems Framework. In: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*. NRS '13, New York, NY, USA, ACM, 2013, pp. 3–9.
- [2] Garg, R., Heena: Study of Text Based Mining. In: *Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*. ACAI '11, New York, NY, USA, ACM, 2011, pp. 5–8.
- [3] Gong, S., Ye, H., Tan, H.: Combining Memory-Based and Model-Based Collaborative Filtering in Recommender System. In: *Proceedings of the 2009 Pacific-Asia Conference on Circuits, Communications and Systems*. PACCS '09, Washington, DC, USA, IEEE Computer Society, 2009, pp. 690–693.
- [4] Ilievski, I., Roy, S.: Personalized News Recommendation Based on Implicit Feedback. In: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*. NRS '13, New York, NY, USA, ACM, 2013, pp. 10–15.
- [5] Robillard, M.P., Maalej, W., Walker, R.J., Zimmermann, T., eds.: *Recommendation Systems in Software Engineering*. Springer, 2014.
- [6] Schein, A.I., Popescul, A., Ungar, L.H., Pennock, D.M.: Methods and Metrics for Cold-start Recommendations. In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '02, New York, NY, USA, ACM, 2002, pp. 253–260.
- [7] Tušer, a kol., A.: *Vydávame školský časopis. Druhé, rozš. vydanie*. Zvolene: Krajská knižnica L. Štúra vo Zvolene, *Televízia Markíza* 2005. 146 s..