

E-NEWSPAPER PERSONALISATION USING SUPPORT VECTOR MACHINE WITH EXPLICIT FEEDBACK

PRIYA CHAUDHARI & RICHA K. MAKHIJANI

Department of Computer & Science, S. S. G. B. COET, Bhusawal, North Maharashtra University, Jalgaon,
Maharashtra, India

ABSTRACT

The number of online journals has increased in recent years owing to the increasing popularity of the Internet. It is important to offer user tools that facilitate faster and more accurate access to articles of interest in digital newspapers. E-Newspaper Personalisation is a very active research field that personalise relevant news to users according to profiles of their visits. News Personalisation systems offer recommendations, usually based on content similarity to previous visits by the user that is the content based approach or on news items visited by similar users that is collaborative filtering. Classifier SVM i.e. Support Vector Machine is used for the text classification purpose. Recommender approach shows news to user as per his choice.

KEYWORDS: Collaborative-Filtering, Content-Based Approach, Explicit Feedback, Personalisation, SVM

INTRODUCTION

The explosive growth of internet and availability of large amount of material about different topics such as science, agriculture, sports etc. led to the design of a system that helps people to find out the information as per they require. The system is personalization system helps to find people to search information as per their interest. The move to personalization is no longer an option, but a necessity. Several challenges however come into place for the personalization procedure to be successful such as scalability, accuracy, evolving user interests, data collection and preprocessing etc. Personalisation system is as content-filtering and collaborative filtering systems. Collaborative filtering methods are based on collecting and analyzing a large amount of information on user's preferences and predicting what users will like based on their similarity to other users. Content-based filtering methods are based on information about and characteristics of the items that are going to be recommended.

Newspaper Personalisation system personalise news to the user from news database. For this purpose news classification, text classification and categorization are also necessary. So for this purpose classifier such as SVM is used as in [6]. SVM i.e. Support Vector Machine is proposed by Vapnik. Support Vector Machines (SVM) is used for text classification because it is a powerful and reliable tool for text categorization. It has strong dependence on the data used in training. The user sees news as per his choice and also he can give feedback to system called as explicit feedback about the news.

PROPOSED SYSTEM

We propose E-Newspaper Personalisation system using which text classification and news personalisation is possible to the user as per his choice. System obtains news obtained from various providers and distributes personalized newspapers to subscribed users. The architecture of E-Newspaper Personalisation System is as in Figure 1. The system consists of five layers: Aggregator, Content Manager, Personalization, Content Delivery Services, and System Management Tools.

- The main responsibility of the aggregator is to check the content providers for new news items, download them and store the new aggregated items in system.

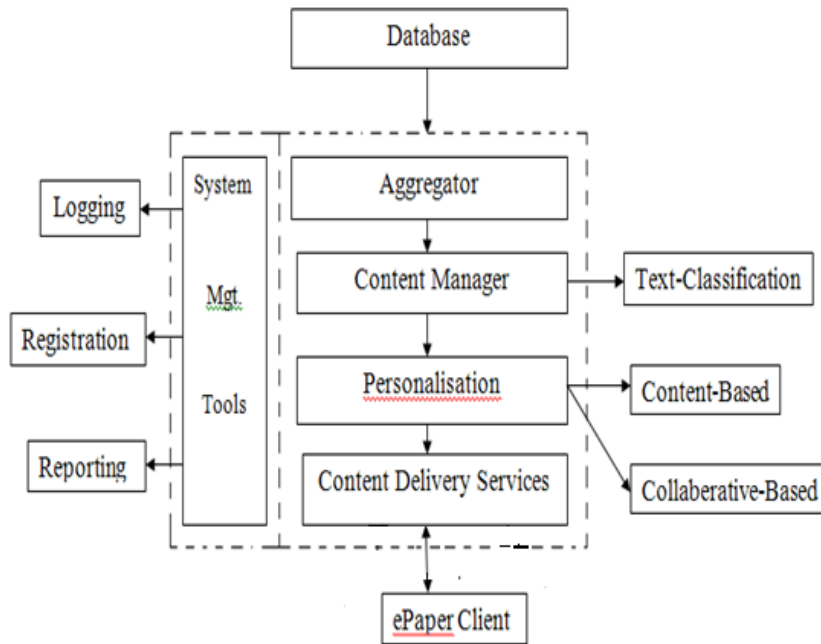


Figure 1: E-Newspaper Personalisation System

- The Content Manager processes the content of each news item received from the aggregator, and prepares it for personalization and delivery to relevant users.
- The Personalisation applies content-based and collaborative filtering.
- The Content Delivery Services interact with the Personalization layer, submits requests for personalized news from users, and sends the ranked news items it receives to the user. It also receives feedback from the user and sends this data to the Personalisation layer.
- The System Management Tools layer provides standard system tools such as logging and reporting, as well as special tools for the News personalisation application.
- The ePaper client interacts with the Content Delivery Services layer for receiving data. The data sent to the subscribed user includes the news of that user's choice.

So this E-Newspaper Personalisation system will provide the user news of his choice through these layers. The aggregator takes the news from the news database and performs the news classification. Content based filtering is the providing news to user as per his choice and collaborative filtering determines the news read by the maximum number of the users and this is done by content manager. Content delivery services take the choice of user and send it to the layer personalization and then accept corresponding news from personalization and display for the user. The system management tools used are logging, registration and reporting.

Database

The news are displayed for the user as per his choice by aggregating the news from the database. So for this purpose a database has been created which contains latest news from different nine categories such as football, cricket, entertainment, technical, automobile, education, hockey, Indian business and environment. The news taken are latest news from month January in 2013. The database is then classified using SVM. We also use 20 Newsgroup dataset. The 20

Newsgroups dataset is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups as given in [3], [4].

IMPLEMENTATION

The E-Newspaper Personalisation system provides personalisation service. Personalisation is providing data to the user as per his choice. As per user is having subscription they also give their choice regarding news that which types of news he want to read. As per user's choice the news is displayed for user.

SVM Algorithm

The personalization system personalise the news to the user as per his choice. This personalization system requires two steps to complete this personalization of news to user. The steps are as below:

- Database Preparation
- Personalisation of news to the user

The database preparation is the classification of news from news database into corresponding categories. The news database consists of news from different categories and that news are classified into corresponding category using SVM classifier as in [5]. SVM i.e. Support Vector Machine is used for news classification purpose.

For multi-classes we might want to apply successively 2-classes learning processes as in [7]. The steps in **database preparation** using SVM are as below:

For Each Category Repeat Following Steps

- Create one dictionary for each category.
- Store all words related to the category in dictionary of that category.
- Svmclass is used to train SVM classifier that is used for news classification.

```
[xsup, w, b, pos, timeps, alpha, obj] = svmclass (x, y, c, lambda, kernel, kerneloption, verbose, span, alphainit)
```

Support vector machine is used for classification. This routine classifies the training set with a support vector machine using quadratic programming algorithm.

- SVM takes each news from news database.
- Calculate the number of terms in news and check that each word in that news related to which category.
- The category to which more number of words related is the category of that news.
- Repeat 3) to 5) steps for each news.
- Svmval is used for testing that the classification is done properly or not i.e. whether the expected output is equal to actual output or not.

```
[y,y1,y2] = svmval (x, xsup, w,b, kernel, kerneloption, span, framematrix, vector, dual)
```

svmval computes the prediction of a support vector machine using the kernel function and its parameter for classification or regression.

- The accuracy of the system regarding displaying news i.e. how accurately the system displays news for the user as per his choice is calculated as :

$$\text{Accuracy} = [\text{Accuracy } 100 - (100 * \text{FP}) / \text{size (Featest, 2) }]$$

Where FP = sum (Exp_out & out)

The Steps for Personalization of News to the User Are as Below

- Select user.
- According to user selection system retrieves the user profile which contains the categories of user's choice.
- System retrieves the news from classified news database as per given in user profile.
- System Produces Personalized Newspaper.
- System Presents the Personalized Newspaper to the User. The news are displayed for the user.

The news classification is done using SVM. SVM helps for database preparation. SVM accepts each news and checks that the news related to which category.

The news database is provided with the dictionary of each category in news database which consist of maximum of all words related to that category. SVM checks each word of each news related to which category using the dictionary.

The category to which maximum number of words in that news is taken as the category of that news and the news is assigned to that category. This is done for each news. After this the news personalisation is done. The user selects his profile as per he has created his profile.

The profile consists of his choice. The system takes his choice and prepares a newspaper by extracting the news from news database. And after that he displays that newspaper for the user.

In this way the news from news database are first of all classified into different categories using SVM classifier and then the news as per user choice is retrieved from corresponding category from classified news database and then the news is displayed as per user's choice by the news personalization system.

PROPOSED FRAMEWORK

The E-Newspaper Personalisation system personalise news to user as per his choice from news online database. When the user wants to see news he has to first select database which is the news database from which the system retrieves news.

User creates his own profile using 'Add new user Profile' option. As per given in figure 2. there is one button 'select database.' After this the user will select his own profile using 'select user profile'.

As per given in figure 2 the user profile of 'Rita Chaudhari' is selected. Rita Chaudhari is having interest in categories 'Entertainment', 'Education' and 'Environment'. So, after the selection of user profile the system prepare a newspaper as per user choice by retrieving news from news database and display that newspaper for user as per given in Figure 2.

If user likes any news then he will give feedback for that news by selecting that news. That selected news is displayed as highlighted news of that day for next users. It is possible to update the news database by posting and classifying new news into it.

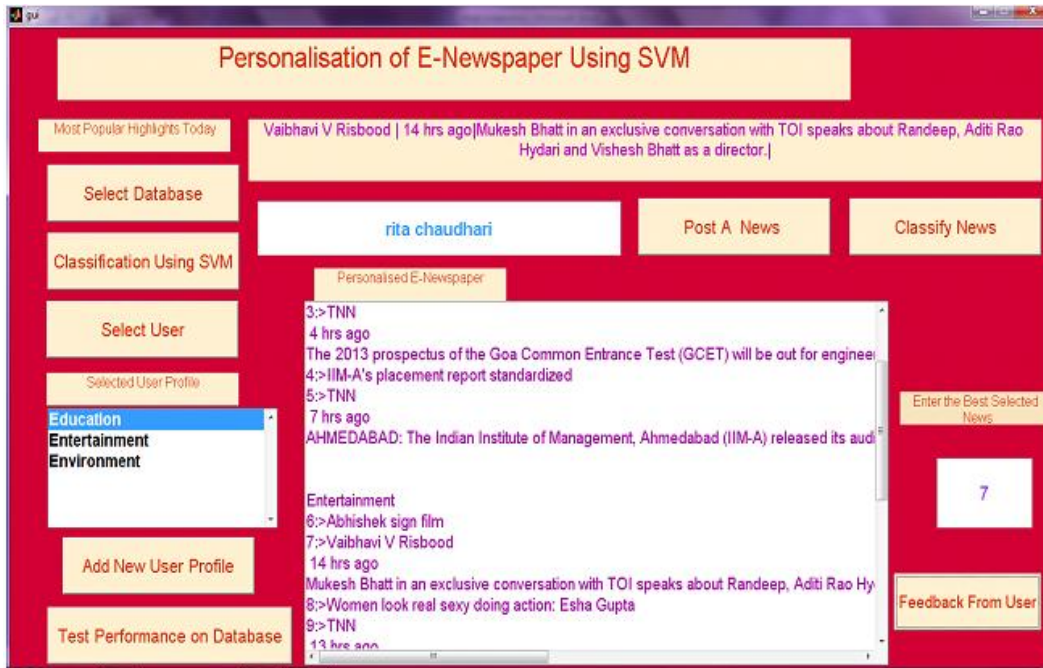


Figure 2: Newspaper Display for User ‘Rita Chaudhari’ and She Prefer 7th News

EXPERIMENTAL EVALUATION

Test Performance Evaluation for Online Database

The online database consists of news from different nine categories such as business, automobile, cricket etc.. SVM classifies that news properly into these nine categories using the dictionary of each category and personalisation system personalises news to user as per his choice. The accuracy provided by SVM for each and every category of online news database is as per given in below Table 1.

Table 1: % Accuracy for Online Database

Name of Category	% Accuracy
Business	99.99
Automobile	100
Cricket	100
Education	100
Entertainment	100
Environment	99.99
Football	100
Hockey	100
Technical	100

As per given in Table 1 accuracy for ‘business’ and ‘Environment’ category is 99.99%. i.e.99.99% accurately the news from news database are classified in category ‘business’ and ‘Environment’. The accuracy of classification for category ‘automobile’ is 100%. For remaining categories of online news database the accuracy is 100%. The accuracy shows how accurately the news from online database are classified among corresponding category which is done using SVM classifier.

Test Performance Evaluation for 20 Newsgroup Dataset

The 20 Newsgroup Dataset is a news database consisting of news from 20 different categories. The accuracy of 20 Newsgroup Dataset is calculated by performing the news classification using SVM. The accuracy table for 20 Newsgroup Dataset is as shown below in Table 2.

Table 2: % Accuracy for Online Database

Name of Category	Accuracy
alt.atheism	99.3897
comp.graphics	99.2562
comp.os.ms-windows.misc	99.2371
comp.sys.ibm.pc.hardware	99.2562
comp.sys.mac.hardware	99.2752
comp.windows.x	99.2180
misc.forsale	99.2562
rec.autos	99.2562
rec.motorcycles	99.2371
rec.sport.baseball	99.2371
rec.sport.hockey	99.2371
sci.crypt	99.2562
sci.electronics	99.2371
sci.med	99.2562
sci.space	99.2371
soc.religion.christian	99.2371
talk.politics.guns	99.3134
talk.politics.mideast	99.2752
talk.politics.misc	99.4087
talk.religion.misc	99.5232

As shown in Table 2 the accuracy values for all categories of 20 Newsgroup Dataset. Accuracy is defined as the agreement between a measured quantity and the true value of that quantity. SVM is applied to each news and the classification of news among 20 categories is done. As per given in table there are accuracy value for category 'alt.atheism' is 99.39%, for comp.graphics its 99.25% and for comp.os.ms-windows.misc 99.23% etc. So these are accuracy values that indicate how accurately the classification of news among categories using SVM is done.

CONCLUSIONS

The E-Newspaper Personalisation System personalise the news to the user as per that user's choice. When any user wants to read news of particular category, he has to search that news. So we propose the E-Newspaper Personalisation System that provides this facility. When subscribed user gets logged in, the personalization system provides news to that user of his choice from the news database as per user's choice is known to the system based on Content-based approach. Based on the collaborative approach it is decided that how many users reading particular news. The SVM i.e. Support vector Machine classifier is used for the text classification purpose. It can operate even in fairly large feature sets as the goal is to measure the margin of separation of the data rather than matches on features. The news is assigned to the user as per his choice and the time of the user is also saved as per he has not to search for the news.

ACKNOWLEDGEMENTS

We would like to thank my teachers in S.S.G.B. COET, Bhusawal, Maharashtra, India. And I specially thankful to my co-author Prof. Richa K. Makhijani.

REFERENCES

1. Sergio Cleger-Tamayo , Juan M. Fernández-Luna , Juan F. Huete "Knowledge-Base System- 'Top-N news recommendations in digital newspapers' ", the research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018) and the Junta de Andalucía, Page 180-189, 2011
2. Andrew Messenger, Jon Whittle, "Recommendation-Based User Generated Comments In Social Media" , IEEE

International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, Page No.505-508, 2011

3. Ioannis Antonellis, Christos Bouras, Vassilis Pouloupoulos, “Scalable Text Classification as a tool for Personalisation”, Pearch Academic Computer Technology Institute N Kazantzaki, University Campus, GR-26500 patras Greece, Page No. 1-3, 18-20.
4. Ming-Wie Chang, Lev Ratinov, Dan Rov, and Vivek Srikumar, “Importance of Semantic Representation : Dateless Classification”, Association for Advancement of Artificial Intelligence.
5. Asa Ben Hur, Jason Weston ,”User’s Guide to Support Vector Machine”, Page No. 1-3
6. Dustin Bosswel, “Introduction to Support Vector Machine”, August 6, 2002, Page No. 1-7
7. Shantanu Godbole, Sunita Sarvagi, Soumen Chakrabarti, “Scaling multi-class Support Vector Machine using Interclass Confusion”, SIGKDD’02 Edmonton, Alberta, Canada, ACM 2002.

