Chapter 7

Personalized Content Recommendation Engine for Web Publishing Services Using Textmining and Predictive Analytics

Başar Öztayşi
Istanbul Technical University, Turkey

Ahmet Tezcan Tekin
Kontra Digital, Turkey

Cansu Özdkicioğlu
Kontra Digital, Turkey

Kerim Caner Tümkaya
Kontra Digital, Turkey

ABSTRACT

Recommendation systems have become very important especially for internet based business such as e-commerce and web publishing. While content based filtering and collaborative filtering are most commonly used groups in recommendation systems there are still researches for new approaches. In this study, a personalized recommendation system based on text mining and predictive analytics is proposed for a real world web publishing company. The approach given in this chapter first preprocesses existing web contents, integrate the structured data with history of a specific user and create an extended TDM for the user. Then this data is used for prediction of the users interest in new content. In order to reach that point, SVM, K-NN and Naïve Bayesian methods are used. Finally, the best performing method is used for determining the interest level of the user in a new content. Based on the forecasted interest levels the system recommends among the alternatives.

INTRODUCTION

Web publishing service includes building and uploading websites, updating the associated webpages, and posting contents to these webpages online. The content meant for web publishing is composed of mainly text, videos, and digital images. Web publishing service providers must hold a web server, a web publishing software, and an Internet connection to carry out publishing process.

DOI: 10.4018/978-1-5225-2148-8.ch007
The main revenue model of web publishers is the advertisement-based model which means providing contents with advertisements and getting fees from advertisers. In this model, advertisers are charged for the number of times readers view or click on the advertisements. The first one is known as “Cost for impression” and the second one is known as “Cost Per Click”. Thus, for publishers the total amount of time the readers spend on the website is the key point to increase the revenues. If the readers spend more time in website, this situation affects the revenue in directly proportional.

As a result, web publishers search for new ways to keep the visitors in the web site. Especially content based websites are making efficient content recommendations to increase current visitors. The most common way for content recommendation is to provide a web part on the page and listing the titles of most popular contents. The basic assumption behind the idea is; the content that have attracted high attention from all visitors will probably get the attention of the current visitor. However, it is obvious that the visitors may have different interests. Thus, systems can be modeled for better recommendations.

The study is conducted in association with Kontra Digital Services (KDS) which is established in 2012 and works as value added services provider for GSM carriers and web publisher since it was founded. The study is conducted for a web portal entitled Gazetemsi, which is a comprehensive web publishing service. Gazetemsi is involved with news on sustainable living, physical and mental health, sports, arts and crafts, culture, entertainment, technology, gaming, automotive, fashion, food, travel and daily news. KDS brings unique news to the readers in order to help them live more fulfilling lives, and feed their sense of wonder, instead of using the internet just for killing time. KDS is differentiated from its competitors by providing “beneficial content with entertaining aspects”. To achieve this, KDS crawls the popular web sites, makes additional researches, merges the information with editors’ knowledge and creativity, and offers contents worth reading or watching. In a nutshell, KDS is in the systematical process of publishing original contents on the website to unique users.

The aim of this study is to search for an efficient content recommendation systems in the web publishing content. The presented approach integrated text mining and predictive analytics to construct a personalized recommendation engine. The paper also provides a real world example for Turkey containing predictive modeling technique such as K-NN, Support Vector Machine and Naïve Bayesian. The rest of the chapter is as follows: in the second section a brief overview of recommendation systems are given. In the third section, after introducing the basics of text mining and predictive analytics the personalized recommendation system is resented with a real world example. After giving, results and recommendations in forth section, future research suggestions are listed. Finally, in the last section the conclusion takes place.

BACKGROUND

Recommendation Systems have been widely investigated in industrial, academic, and educational fields. There are many approaches in literature. In the literature, the best known approaches are content based filtering and collaborative filtering (CBF). In CBF, items are grouped in specific properties. When users register a system firstly user profile is created for each of them. The profile is defined by items which are examined, liked or bought by users before. Based on this user profile, list of item recommendation is defined. Schein & Popescul (2002) underline that content based filtering ignore the preferences of the other users. Each recommendation engine has its own advantages and disadvantages when compared...