Nov 1st, 2:30 PM - 2:45 PM

Using Machine Learning Techniques to Customize the User's Profile, Helps Intelligent TV Decoder's Design

Alketa Hyso
Ismail Qemali University, alketa.hyso@univlora.edu.al

Roneda Mucaj
Ismail Qemali University, roneda.mucaj@univlora.edu.al

Using Machine Learning Techniques to Customize the User's Profile, Helps Intelligent TV Decoder's Design

Alketa Hyso[1], Roneda Mucaj[2]
[12] Computer Science Department, "Ismail Qemali" University, Vlore, Albania
{alketa.hyso[1], roneda.mucaj[2]}@univlora.edu.al

Abstract. In today's society due to the increase of the quantity of information is becoming more difficult to find the information we search. "Data mining" offers us the most important methods and techniques in data analysis. Through this work, we aim to study the several data mining techniques, methods and applications in specific areas. We experiment with an "open software" WEKA, to perform some data analysis, presenting the reliability and advantages of data mining classification technique. We use the decision trees technique to achieve the task of classification, to customize user profiles based on their requirements and needs. This paper presents also how machine learning methods can be integrated with agent technology in building more intelligent agents. Using machine learning techniques makes it possible to develop agents able to learn from and adapt to their environment. So a TV decoder can be adapted to the demands of TV viewers. If the decoder initially trained by the demands and needs of viewers, it can display intelligent behavior, suggesting viewers, according to the profile created for each one, shows and movies. The paper concludes with our contributions concerning the application of data mining techniques to customize services according to the requirements and needs of users.

Keywords: machine learning, data mining, data analysis, WEKA

1 Introduction

Internet, computer networks, data collection systems, enable us to collect and preserve large amounts of data. Scientific research, and different aspects of social life have accumulated large amounts of data, that bring difficulties in extracting useful information. Analysis of the relationships to the data to find useful information there is a need. But computer technology and techniques of data collection can not help us to analyze and understand the data. A few years ago, for the extraction of knowledge, We used the usual methods, such as analysis, filtering, and comparison. Based on them, we take out knowledge and create rules.

Today, at a time when the majority of data are stored in data warehouses, old methods do not work anymore. In response to these challenges was developed data mining technology. Data mining enables the transformation of large amounts of data in useful information, readily understandable by humans, and widely applicable to various applications. Data mining technology emerged as a technology-oriented application. Information and knowledge can be widely used in various applications, including business management, production control, market analysis, engineering design and scientific exploration.

Data—mining technologies have been around for decades, without moving significantly beyond the domain of computer scientists, statisticians, and hard-core business analysts. New data mining technologies developed can support personalization. Personalization technologies treated as a process-oriented perspective [1]. Are developed systems which construct personal profiles based on customers' transactional histories. The system uses data mining techniques to discover a set of rules describing customers' behavior and supports human experts in validating the rules [2]. New improvements are made in data mining and its application to personalization in E-commerce[3]. Personalization is what merchants and publishers want to do to tailor the Web site or advertisement and product promotion to

a customer based on his past behavior and inference from other like-minded people examine [4]. The need for new marketing strategies such as one-to one marketing and customer relationship management brings personalized recommendation that helps customers find the products they would like to purchase by producing a list of recommended products for each given customer [5].
In this paper we will treat personalization technology that gives us, information about the person, so we can predict his attitudes and behaviors. Personalization takes place by making recommendations to the user. A case study will be analyzed to profile the behavior outlined in the viewers about television programs. Furthermore, the model obtained after analysis of the data is implemented on a TV decoder. This model provides viewers with recommendations on TV programs that are classified in the group of his favorites. Finally, we summarize our contributions concerning the application of data mining techniques to customize services according to the requirements and needs of users, extracting knowledge and helping decision making, as well as bringing suggestions for future research.

2 User's Profile

Personalization technology provides information about persons to predict their attitudes and behaviors. "A user profile is a description of an individual user and contains the most important or interesting facts about him. It is necessary to build user profile, because users have different preferences, interests, and purposes. These differences form the basis of personalization. There are several types of information that can be stored in a user profile [6]:
Personal information. Includes age, gender, city, state, etc. Interests of a user, are essential part that makes personalization. Interests may be related to hobbies, work issues, etc. Behavior, is the type of information collected in a non-obvious way. The goals of the user, are important to detect. Information can be obtained implicitly and explicitly, through surveys of their behavior, based on geographic location, social groups, etc. Models can be detected using techniques such as machine learning and data mining, turning the collected information into a value.

3 Recommendation

Assume that we have a system that monitors the preferences of a person watches TV, in order to recommend to him, other performances, that he may please. The viewer is open to suggestions. Recommendations can be entered here. A variety of recommendation techniques has been developed [7]. Recommendations can be made according to the following categories:
Recommendation by the program: recommendation is based on the program itself. This approach uses a content-based filter. There are recommended programs with similar content. One way to do this is through the use of tag. Another way is through machine learning techniques such as neural networks or decision trees.
Recommendations based on social group, part of which the user is: recommendation of programs is based on similar audiences' preference. This approach uses collaborative filtering. The viewer is suggested programs, that social group prefer. The idea behind this is that, the viewer receives recommendations from people with similar tastes. Collaborative filtering has been used in a number of different applications such as recommending web pages, articles and products.
Personalized recommendations. Recommended programs are based on past behavior of viewers. This approach combines based filtering by content and collaborative filtering. Recommendations are based on the programs that the user has previously preferred and other programs favored by other viewers, with whom they have similar tastes and preferences.
An adaptive decoder automatically improves the organization of programs according to the preferences of viewers by learning from the viewer preferences data. Reorganization of the programs may vary in consistency. Displaying proper program to a particular viewer can be seen as a prediction of behavior, prediction of the behavior of a viewer. Although personalization can create advantages for many viewers, personalization technology should be applied with caution. There are several problems associated with personalization, which should be taken into consideration when implementing such

techniques [9]. One concern is that personalized filters limit what appears to the viewer, influencing so, the way we think and learn. Presenting things that we know, things that we are interested in at the moment can prevent us from treating different perspectives from ours. This can strongly influence on personal development.

4 Case Study

In today's society, where people are offered by private television stations to multiple programs, it is increasingly difficult for viewers to determine the program, they want to watch. In order to provide a viewer to the right information, it is important to collect information about viewer, processing it, and suggesting further information, that he would be interested. Therefore, it is necessary to collect specific information about him, and to create a profile of viewer's behavior in relation to television programs. Decision trees, is a method that we have chosen to achieve the task of classification. We address the use of decision trees to gather information for viewers based on their preferences.

4.1 Data Preparation

In our study we have taken into consideration 12 examples. Films are categorized as follows: comedy, history with doctors, lawyers, or action movie with guns. In the table below we have collected information on the preferences of the viewer on TV movies. A taxonomy of the movies can be seen as follows:

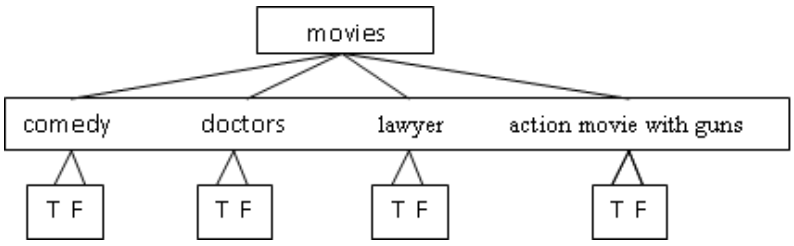

Fig. 1. A taxonomy of the movies

We will use this set of data, Table 1, to find the values of attribute "like", (ie, to predict what movie would have liked the viewer based on the attributes of the film on TV). So, we use an open software WEKA, where we apply the decision trees technique.

Table 1. Examples of training. Data tuple.

Example	comedy	doctor	lawyer	action movie with guns	*Like*
e_1	false	true	false	false	false
e_2	true	false	true	false	true
e_3	false	false	true	true	true
e_4	false	false	true	false	false
e_5	false	false	false	true	false
e_6	true	false	false	true	false
e_7	true	false	false	false	true
e_8	false	true	true	true	true
e_9	false	true	true	false	false
e_{10}	true	true	true	false	true
e_{11}	true	true	false	true	false
e_{12}	false	false	false	false	false

WEKA provide a uniform interface to many different learning algorithms, tools for preprocessing the data, called filters, and for evaluating the result of learning schemes on any given dataset. So, we can preprocess a dataset, feed it into a learning scheme, and analyze the resulting classifier and its performance. Another way of using WEKA is to apply several learners and compare their performance in order to choose one for prediction [10]. WEKA systemsupports three types of data file, files imported from local database files, site database, or databases. We use the data format called ARFF (Attribute-Relation File Format). ARFF data format for our study case is as follows:

@relation preferenceTV
@attribute examples {e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11, e12}
@attribute Comedy{True, False}
@attribute Doctor {True, False}
@attribute Lawyer {True, False}
@attribute Guns {True, False}
@attribute Like{True, False}
@data
e1 ,False ,True ,False ,False ,False
e2 ,True ,False ,True ,False ,True
e3 ,False ,False ,True ,True , True
e4 ,False ,False ,True ,False ,False
e5 ,False ,False ,False ,True ,False
e6 ,True ,False ,False ,True , False
e7 ,True ,False ,False ,False , True
e8 ,False ,True ,True ,True , True
e9 ,False ,True ,True ,False , False
e10 ,True ,True ,True ,False , True
e11 ,True ,True ,False ,True , False
e12 ,False ,False ,False ,False , False
The ARFF file can be divided into two parts:
(1) Head information: @ relation defines the name of data set, equivalent to the name of a database table. @ Attribute defines an attribute of the data set; it contains the attribute name and possible values of attribute or the attribute type.
(2) Data information : @ defines the start of data set records, followed by the entire data set records.

Levels of data mining process are:

- Data input layer: This is the preparation phase of the whole data mining.
- Data mining layer: This includes preprocessing, classification, clustering and other functions;
- Model evaluation layer: It takes model assessment on the result of data mining, analyzes the results of data mining.
- Visualization layer: It achieves data visualization, mining process visualization, and mining result visualization.
- Storage layer: It uses a specific format to store the mining results.

4.2 The Usual Data Mining Method - Classification

Data mining combines artificial intelligence methods and machine learning with traditional methods of statistical analysis, mathematical methods, scientific visualization. Methods aim to find useful patterns in the data and gain knowledge. Categorization reveals a number of common characteristics of data in a database, and the classification model divides data into different classes, in order to classify data elements from the database in a given category, providing target discrete variables [11].

Classification is a data mining task that categorizes a set of cases in a database into different classes according to a classification model. For this task, a model set (i.e., a set of cases whose class labels are known) is first analyzed and a classification model is constructed based on the features available in the data of the model set. Such a classification model is then used to categorize a score set (i.e., a set of cases whose class labels are unknown) [12]. The most popular classification method is the decision tree induction which builds a decision tree and performs classification on the given data using it.

4.3 Extraction of the decision tree, that recommends the movie to the viewer

Based on WEKA, after training the data set in Table 1 with Decision Tree algorithm J.48, to learn the decision tree, we arrive at a decision tree as in Figure 2; Based on the decision tree model, we can easily draw some rules, that can be used by decoder to recommend the viewer. In the tree above we can read starting from the root and descending towards each leaf:

rule 1.
if subject is (“lawyer” and “comedy”)
then target “Like” is TRUE;

rule 2.
if subject is not (“lawyer”)
then target “Like” is FALSE;

rule 3.
if subject is (“lawyer” and “ not comedy” and “guns”)
then target “Like” is TRUE;

rule 4.
if subject is (“lawyer” and “not comedy” and “ not guns”)
then target “Like” is False;

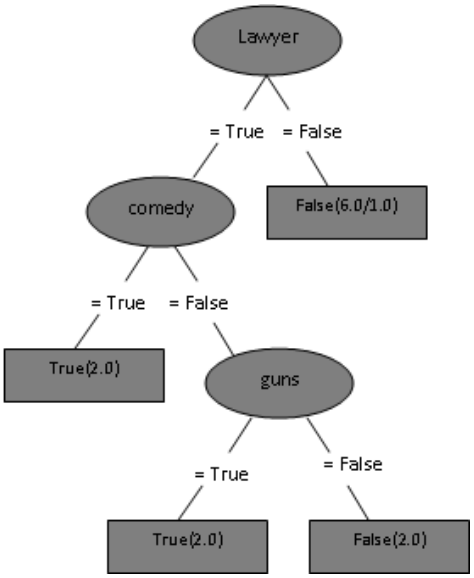

Fig. 2. Visualization of the decision tree preferences for TV movie viewer

We evaluate the error in the decision tree model. This error is 0.5. This is because the set of examples, we have taken into consideration is small. If data set grows and we take into consideration more

attributes, the decision tree will clearly present the viewer preferences. The overall flow of recommendation is presented in figure 3.

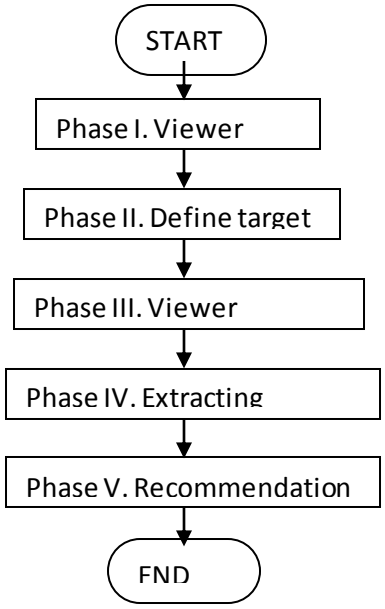

Fig. 3. The overall flow of recomandation

4 Conclusions

Through this paper we presented a data mining methods - classification. We implemented the technique of decision trees, as a step to enable the recommendation according to the viewer profile. Data mining is compelling tools that can be used not only for successful extraction of knowledge but also in decision support and predictive analysis, helping to adapt our household equipment and self-management of information systems.

References

1. Adomavicius, G., Tuzhilin, A. Personalization technologies: a process-oriented perspective. Communications of the ACM - The digital society. Volume 48 Issue 10, Pages 83-90. October 2005
2. Adomavicius, G., Tuzhilin, A. Using data mining methods to build customer profiles. Computer (Volume:34 , Issue: 2). ISSN :0018-9162. Feb 2001 Page(s): 74 – 82,
3. Sanchez, M., Moreno, N. M., Segrera, S. Framework for the development of a personalised recommender system with integrated web-mining functionalities. International Journal of Computer Applications in Technology. Volume 33 Issue 4, Pages 312-319. January 2009
4. Watson, T, J.. Data mining and personalization technologies. Database Systems for Advanced Applications, 1999. Proceedings., 6th International Conference. Page(s): 6 – 13. 1999.
5. Cho, Y. H., Kim, J. K., Kim, S. H. A personalized recommender system based on web usage mining and decision tree iduction. Expert Systems with Applications 23page: 329–342. 2002.
6. Schiaffino, S. and Amandi, A. (2009) Intelligent User Profiling. Available from URL: http://www.exa.unicen.edu.ar/catedras/knowmanage/apuntes/56400193.pdf [Accessed 20 Septembre 2012]
7. Lawrence, R. D., Almasi, G. S., Kotlyar, V., Viveros, M. S., & Duri, S. S. (2001). Personalization of supermarket product recommendations. Data Mining and Knowledge Discovery, 5(1–2), 11–32.

8. Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use Proceedings of the 1995 ACM Conference on Human Factors in Computing Systems (pp. 194–201).
9. Treiblmaier, H., Madlberger, M., Knotzer, N. and Pollach, I. (2004) Evaluating Personalization and Customization from an Ethical Point of View: An Empirical Study. Available from URL: https://springerlink3.metapress.com/content/ h116661574707141/resource-secured/?target=fulltext.pdf&sid=f0gtf5bbpfjitjqr1i30awjd&sh= www.springerlink.com [Accessed 28 Septembre 2012].
10. Witten, H.Ian., Frank,E. WEKA, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.(2000). Morgan Kaufmann Publishers
11. Tan, P., Steinbach, M. and Kumar, V. (2005) Introduction to Data Mining. Boston: Addison Wesley. ISBN: 978-0321321367.
12. Cho, Y. H., Kim, J. K., Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree iduction. Expert Systems with Applications 23page: 329–342.