

Web mining homework

[Education](#), [Homeschooling](#)



Abstract

In this paper, we introduce a web data mining solution to e-learning system to discover hidden patterns strategies from their learners and web data, describe a personalized recommender system that uses web mining techniques for recommending a student which (next) links to visit within an adaptable e-learning system, propose a new framework based on data mining technology for building a Web-page recommender system, and demonstrate how data mining technology can be effectively applied in an e-learning environment.

Keywords--Data mining; web log,; e-learning; recommender readily interpreted by the analyst. A virtual e-learning framework is proposed, and how to enhance e-learning through web data mining is discussed.

Introduction

With the rapid development of the World Wide Web, Web data mining has been extensively used in the past for analyzing huge collections of data, and is currently being applied to a variety of domains [1]. In the recent years, e-learning is becoming common practice and widespread in China.

With the development of e-Learning, massive amounts of learning courses are available on the e-Learning system. When entering e-Learning System, the learners are unable to know where to begin to learn with various courses. Therefore, learners waste a lot of time on e-Learning system, but don't get the effective learning result. It is very difficult and time consuming for educators to thoroughly track and assess all the activities performed by all learners.

In order to overcome such a problem, the recommender learning system is required. Recommender systems are used on many web sites to help users find interesting items [2], then predict a user's preference and suggest items by analyzing the past preference information of users, e-learning system is applied on the basis of the method. The user's learning route is given and then provides the relevant learners useful messages through dynamically searching for the appropriate learning profile.

This paper recommends learners the studying activities or learning profile through the technology of Web Mining with the purpose of helping they adopt a proper learning profile, we describe a framework that aims at solution to e-learning to discover the hidden insight of learning profile and web data. We demonstrate how data mining technology can be effectively applied in an e-learning environment. The framework we propose takes the results of the data mining process as input, and converts these results into actionable knowledge, by enriching them with information that can be

The route where the learner browses through the web pages will be noted down in Web log, carries on the technology of Web mining through Learning Profile and Web log, and analyzes from the materials related to association rule. It can be found the best learning profile from this information. These learning profiles combine with the Agent and put them on the learning website. Furthermore, the Agent recommends the function of learning profiles on learning website. Therefore, the learner will acquire a better learning profile.

This chapter briefly illustrates the relevant contents including: e-Learning, Learning Profile, Agent, Web Data mining and Association rule. A. E-learning
<https://assignbuster.com/web-mining-homework/>

E-learning is the online delivery of information for purposes of education, training, or knowledge management. In the Information age skills and knowledge need to be continually updated and refreshed to keep up with today's fastpaced study environment. E-learning is also growing as a delivery method for information in the education field and is becoming a major learning activity. It is a Web-enabled system that makes knowledge accessible to those who need it.

They can learn anytime and anywhere. E-learning can be useful both as an environment for facilitating learning at schools and as an environment for efficient and effective corporate training [3]. B. A Glance at Web Data Web usage mining performs mining on web data, particularly data stored in logs managed by the web servers. All accesses to a web site or a web-based application are tracked by the web server in a log containing chronologically ordered transactions indicating that a given URL was requested at a given time from a given machine using a given web client (i. e. browser).

As shown in table 1, Web log contains the website “ hit” information, such as visitor's IP address, date and time, required pages, and status code indicating.

Support, as usefulness of a rule, describes the proportion of transactions that contain both items A and B, and confidence, as validity of a rule, describes the proportion of transactions containing item B among the transactions containing item A. The association rules that satisfy user specified minimum support threshold (minSup) and minimum confidence threshold (minCon) are called strong association rules. D. Web Mining for E-learning Learning profile

help learner to keep a record of their current knowledge and understanding of e-learning and elearning activities.

Web mining is the application of data mining techniques to discover meaningful patterns, profiles, and trends from both the content and usage of Web sites. Web usage mining performs mining on web data, particularly data stored in logs managed by the web servers. The web log provides a raw trace of the learners' navigation and activities on the site. In order to process these log entries and extract valuable patterns that could be used to enhance the learning system or help in the learning evaluation, a significant cleaning and transformation phase needs to take place so as to prepare the information for data mining algorithms [6].

Web server log files of current common web servers contain insufficient data upon which to base thorough analysis. The data we use to construct our recommended system is based on association rules.

Recommendation Using Association Rules

One of the best-known examples of data mining in recommender systems is the discovery of association rules, or item-to-item correlations [7]. Association rules have been used for many years in merchandising, both to analyze patterns of preference across products, and to recommend products to consumers based on other products they have selected.

Recommendation using association rules is to predict preference for item k when the user preferred item i and j , by adding confidence of the association rules that have k in the result part and i or j in the condition part [4]. An association rule expresses the relationship that one product is often

purchased along with other products. The number of possible association rules grows exponentially with the number of products in a rule, but constraints on confidence and support, combined with algorithms that build association rules with item sets of n items from rules with $n-1$ item sets, reduce the effective search space.

Association rules can form a very compact representation of preference data that may improve efficiency of storage as well as performance. In its simplest implementation, item-to-item correlation can be used to identify "matching items" for a single item, such as other clothing items that are commonly purchased with a pair of pants. More powerful systems match an entire set of items, such as those in a customer's shopping cart, to identify appropriate items to recommend. The web data is massive since the visitor's every click in the website will leave several records in the tables.

This also allows the website owner to track visitors' behavior details and discover valuable patterns.

C. Data Mining Techniques

The term data mining refers to a broad spectrum of mathematical modeling techniques and software tools that are used to find patterns in data and use these to build models. In this context of recommender applications, the term data mining is used to describe the collection of analysis techniques used to infer recommendation rules or build recommendation models from large data sets.

Recommender systems that incorporate data mining techniques make their recommendations using knowledge learned from the actions and attributes of users. Classical data mining techniques include classification of users, finding associations between different product items or customer behavior,

and clustering of users [4]. 1) Clustering Clustering techniques work by identifying groups of consumers who appear to have similar preferences. Once the clusters are created, averaging the opinions of the other consumers in her cluster can be used to make predictions for an individual.

Some clustering techniques represent each user with partial participation in several clusters. The prediction is then an average across the clusters, weighted by degree of participation. 2) Classification Classifiers are general computational models for assigning a category to an input. The inputs may be vectors of features for the items being classified or data about relationships among the items. The category is a domain-specific classification such as malignant/benign for tumor classification, approve/reject for credit requests, or intruder/authorized for security checks.

One way to build a recommender system using a classifier is to use information about a product and a customer as the input, and to have the output category represent how strongly to recommend the product to the customer. 3) Association Rules Mining Association rule mining is to search for interesting relationships between items by finding items frequently appeared together in the transaction database. If item B appeared frequently when item A appeared, then an association rule is denoted as $A \rightarrow B$ (if A, then B).

The support and confidence are two measures of rule interestingness that reflect

A Visual Web Log Mining Architecture for Personalized E-learning Recommender System

In this section, we present A Visual Web Log Mining Architecture for e-learning recommender to enable personalized, named V-WebLogMiner, which relies on mining and on visualization of Web Services log data captured in elearning environment. The V-WebLogMiner is such a model: with the mining technology and analysis of web logs or other records, the system could find learners' interests and habits. While an old learner is visiting the website, the system will automatically match with the active session and recommend the most relevant hyperlinks what the learner interests. As shown in Figure1, V-WebLogMiner is a multi-layered architecture capable to deal with both Web learner profiles and traditional Web server logs as input data. It maintains three main components: data preprocessing module, Web mining module and recommendation module.) Web Mining Module The Web mining module discovers valuable knowledge assets from the data repository containing learners' personal data by executes the mining algorithms, tracked data of learners' performance and behavior, automatically identify each learner's frequently sequential pages and store them to recommend database. When the learner visit the site next time, hyperlinks of those pages will be added so that the learner could directly link to his individual pages being remembered.

The major component of Web mining module is Web data mining which acts as a conductor controlling and synchronizing every component within the module. The Web data mining module is also responsible for interfacing with the storage. The learning profile evaluation component provide profiling tool

to collect personal data of learner and tracking tool to observe learners' actions including like and dislike information. For personalization applications, we apply rule discovery methods individually to every learner's data.

To discover rules that describe the behavior of individual learner, we use various data mining algorithms, such as Apriori [8] for association rules and CART (Classification and Regression Trees) [9] for classification. 3) Recommendation Module The recommendation module is a recommendations engine; it is in charge of bulk loading data from course database, executing SQL commands against it and provides the list of recommended links to visualization tools.

For the recommendation module, recommendations engine is responsible for the synchronizing process indexing and mapping, is a component for storing and searching recommend assets to be used in the learning process. The recommendation engine considers the active learners in conjunction with the recommended database to provide personalized recommendations, it directly related to the personalization on the website and the development of elearning system. The task of the recommendation engine is to determine the type of the learner online and compute recommendations based on the recent actions of that learner.

The decision is based on the knowledge attained from the recommended database. The recommender engine is activated each time that the learner visits a web page. First, if there are clusters in the recommended database, then the engine has to classify the current learner to determine the most likely cluster. We have to communicate with the engine to know the current

number of pages visited and average knowledge of the learner. Then, we use the centroid minimum distance method [10] for assigning the learner to the cluster whose centroid is closest to that learner.

Finally, we make the recommendation according to the rules in the cluster. So, only the rules of the corresponding cluster are used to match the current web page in order to obtain the current list of recommended links [11]. 4)

The Visualization tools Visualization tools should be used to present implicit and useful knowledge from recommendations engine, Web services usage and composition. Data can be viewed at different levels Figure 1. A visual web mining architecture for Personalized E-learning Recommender System) Data Preprocessing Module The data preprocessing module is set of programs used to prepare data for further processing. For instance: extraction, cleaning, transformation and loading. This module uses Web log files and learner profile files to feed the data repository. The data preparation component is used to parse and transform plain ASCII files produced by a Web server to a standard database format. This component is important to make the architecture independent from the Web server supplier. of granularity and abstractions as patrolled coordinate's graphs [12, 13].

This visual model easily shows the interrelationships and dependencies between different components. Interactively, the model can be used to discover sensitivities and to do approximate optimization, etc. B. The Procedure of the Data is Explained As show in figure 1, the beginning learner, that is to say the earliest one, will study in the e-Learning teaching platform. The course materials of Web studying system come from the

course database. The data of learner's learning profiles may be recorded in the learner profile files and Web log files.

Then next step is to find out the best learning profile from the proceeded data of Web log through web mining to proceed with Association rule and others data mining algorithm. These learning profiles need to be classified—every field has relevant courses and better learning profiles. The recommender engine will offer the list of recommended links when learners study the courses. With the above information and learning profiles, when the future learners study in Web, recommender engine offers related link lists according to recommend database. However, these link lists may not be suitable for all learners.

Therefore, after finishing recommendation every time, there are systems of assessing. The learner ($n + 1$) evaluates the learning profiles that are recommended. Because the profiles analyzed by system may not be perfect, if there are adjustments of evaluation would make the recommendation conform to learners' asks more. These suggestions can help learners navigate better relevant resources and fast recommend the on-line materials, which help learners to select pertinent learning activities to improve their performance based on on-line behavior of successful learners.

Conclusion and Future Work

There are some possible extensions to this work. Research for analyzing learners' past studying pattern will enable to detect an appropriate. Furthermore, it will be an interesting research area to effectively judge session boundaries and to improve the efficiency of algorithms for web data

mining. ACKNOWLEDGMENT The authors gratefully acknowledge the financial subsidy provided by the Xiamen Science and Technology Bureau under 3502Z20077023, 3502Z20077021 and YKJ07013R project.

References

1. D. J. H and, H. Mannila, and P. Smyth. Principles of Data Mining. MIT Press, 2000. J. B. Schafer, J. A. Konstan, and J. Riedl. Recommender systems in ecommerce. In ACM Conference on Electronic Commerce, pages 158166, 1999.
2. Liaw, S. & Hung , H. How Web Technology Can facilitate Learning. Information Systems Management, 2002.
3. Choonho Kim and Juntae Kim, A Recommendation Algorithm Using Multi-Level Association Rules, Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence, p. 524, October 13-17, 2003.
4. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaurmann Publishers, 2000
5. Za ane, O. R. & Luo, J. Towards evaluating learners' behaviour in a web-based distance learning environment. In Proc. of IEEE International Conference on Advanced Learning Technologies (ICALT01), p. 357-360, 2001.
6. Sarwar, B. , Karypis, G. , Konstan, J. A. , & Reidl, J. Item-based Collaborative Filtering Recommendation Algorithms. Proceedings of the Tenth International Conference on World Wide Web, pp. 285 - 295, 2001.

7. R. Agrawal et al. , Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, Calif. , 1996, chap. 12. L. Breiman et al. Classification and Regression Trees, Wadsworth, Belmont, Calif. , 1984.
8. MacQueen, J. B. Some Methods for classification and Analysis of Multivariate Observations. In Proceedings of of 5-th Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281-297.
9. Cristobal Romero, Sebastian Ventura and Jose A. Delgado et al. , Personalized Links Recommendation Based on Data Mining in Adaptive Educational Hypermedia Systems, Creating New Learning Experiences on a Global Scale, 2007, pp. 292-306.
10. Inselberg, A. Multidimensional detective, In IEEE Symposium on Information Visualization, 1997, vol. 00, p. 00-110 .
11. Ware, C. Information Visualization: Perception for Design, Morgan Kaufmann, New York, 2000. [