World Transactions on Engineering and Technology Education
Vol.13, No.3, 2015

Intelligent management of on-line video learning resources supported by Web-mining technology based on the practical application of VOD

Binwen Huang

Hainan Vocational College of Political Science and Law
Hainan, Peoples' Republic of China

ABSTRACT: Faced with the contradictions of an increase in use and low efficiency of on-line learning video resource access, proposed in this study is an intelligent video classification and optimisation algorithm for *video on demand* (VOD). It is based on data mining (DM) technology, and the work demonstrated that it can improve the efficiency of a video on demand system. The results of the study are expected to accelerate the use of video on demand in education and teaching as a result of the use of optimisation. Also it can provide reference and support for on-line open network course construction, such as massive open on-line courses (MOOCs).

INTRODUCTION

Vision is recognised as the major means whereby learners receive information. Video resources allow independent study in an open network environment. At present, video resource construction has emphasised video open classes (on-demand on-line access to a class), micro-lectures (on-line lecture about a specific, defined topic) and MOOCs (a massive open on-line course is a complete course presented on-line providing lectures, tutorials and coursework), but other on-line video-learning resources are developing very quickly. With access to a large number of video resources, learners easily get confused when searching for a resource. Faced with contradictions between video resources and low efficiency in their use, this study is a proposal and discussion of an optimisation algorithm for video material management for VOD (video on demand) applications. As well, this work provides a reference for on-line video learning material management.

PROMINENT PROBLEMS OF VOD

Video on demand, also called interactive TV-on-demand, emerged in the 1990s; it was first television on demand and, then, video on demand. Video on demand is an interactive video system that can offer information material according to a user's demand for many kinds of interactive information services. Nowadays, taking advantage of network technology, video technology and streaming media technology, VOD provides video and text in realising the goal of providing video materials on demand. Learners are able to choose learning content based on their own need.

Video on demand modifies the system where information is transmitted and received passively to an active transmission mode. Massive open on-line courses (MOOCs) are a typical application for VOD in education. Video on demand plays an important pioneering role in the use of on-line open learning materials, such as in MOOCs.

Twenty years on and VOD has had a great impact on education in general, and education reform, such as autonomic learning. As network technology continues to develop rapidly and educational TV has declined, the VOD system could show an obvious advantage of being able to search learning materials quickly and easily. As one of the few media survivals among traditional education technology, VOD is challenged by new problems.

With the rapid rise and wide application of MOOCs in educational teaching, more and more schools are paying greater attention to MOOC innovations and teaching approach. The expansion of user numbers and video materials, as well as the consequent increase of VOD scale and difficulty, has led VOD managers to modify site structures and video materials accordingly, to reflect user interest, number of visits and visit times, in order to meet users' personalised needs.

Based on the above, the study reported in this article explored the question as to how to employ Web data mining to improve VOD system efficiency. Approaches to improving VOD system functions and service quality by Web mining were identified. Meanwhile intelligent, interactive classification systems were investigated.

ADVANTAGE OF WEB DATA MINING

With the fast development of computer networks, there is a flood of material on the Web for VOD. Web/VOD data structures are very large, dynamic and complex. It is high time for data mining (DM) to be applied to VOD data structure and processes.

Data mining is a process of extracting information or knowledge, which cannot be known explicitly and has potential significance, from large quantities of data through computer and database technology [1]. By using various analytical tools, DM discovers relationships between patterns in large amounts of data. The DM supports retrievals and searches from databases; it also involves methods for mic-statistics (maximal information coefficient) or mac-statistics, analysis, organisation and reasoning. As an interdisciplinary sub-field, data mining has incorporated many advanced techniques from other domains, such as databases, pattern recognition, neural networks, information retrieval, artificial intelligence, machine learning, statistics, performance computing, data visualisation, image and video processing, and special data analysis [2][3]. Figure 1 shows the data mining method.

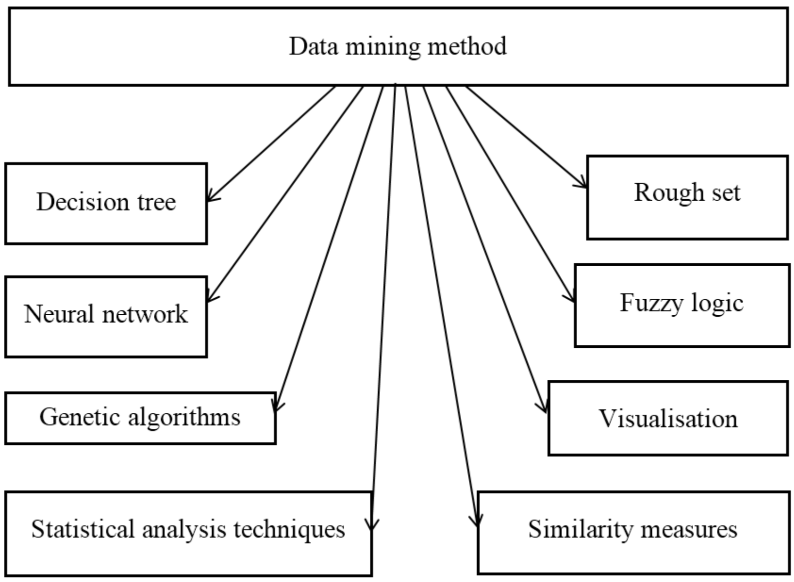

Figure 1: The data mining method.

In this data age, the results from traditional data mining technology can exceed the human ability to comprehend. However, when traditional data mining technology is combined with the Web, learners are better able to find out the useful information that they seek. Through analysis of a large amount of data by exacting, filtering, switching and analysing according to a specific need, decisions and management based on those data become more precise and objective, with efficiency improved.

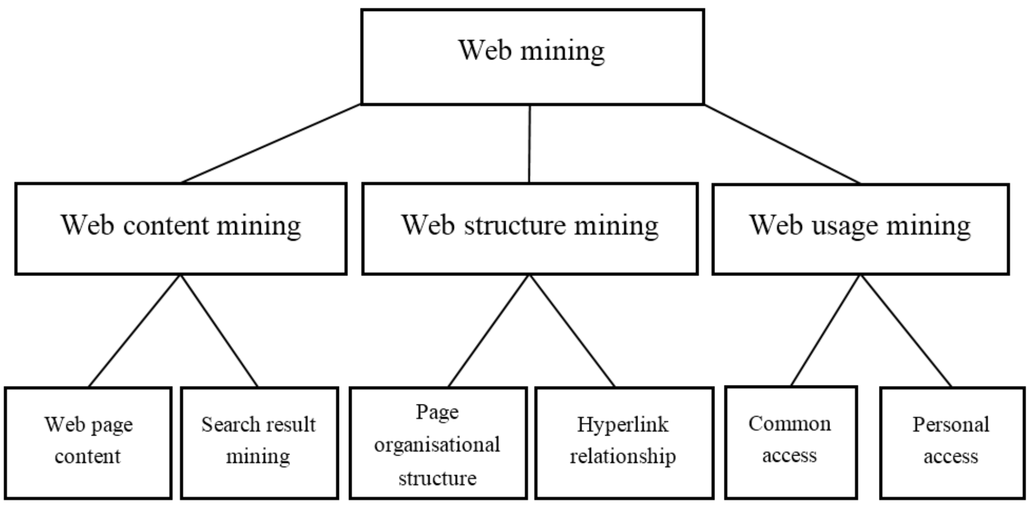

Figure 2: Web mining types.

Web data mining, as shown in Figure 2, also called Web mining, belongs to the emerging field of data mining, and is the application of data mining techniques to discover patterns in data on the World Wide Web. Based on data sources, Web mining can be divided into three types: Web content mining, Web structure mining and Web log mining [4].

WEB DATA CLASSIFICATION AND CLUSTER ANALYSIS

A classification guideline is in the process of being built, to describe data classes and to classify new data. Its major methods include decision tree classification, Bayes classification, back propagating classification, K-nearest neighbour classification and association mining classification. Using classification rules, descriptions of public attributes of one group is applied to a classified visitor [5].

Clustering can be used to organise a large number of visitors who access the Web for information and data in groups, where visitors within a group share strong similar characteristics. This facilitates the development of the service mode and population of the Web log.

Path Analysis and Video Classification

At present, most of the VOD systems adopt as a classification method the user's main retrieval clue, such as action, comedy, science, education. These usually tend to be subjective and unilateral, and cannot properly represent the user's complete thinking. Hence, finding a video in this way is inefficient. To improve the query efficiency, a better classification method has been designed:

- Analyse the access path and visit times;
- Process those data;
- Classify the access path as the user's expected classification;
- Using accumulated data, the highest expectation for video classification will be selected, together with a descending order of classifications.

Nowadays, the sources of data for Web mining mainly come from log files recorded on servers. This records all the user access behaviour during a period. Due to noisy data in Internet logs, user access path data are collected, transformed to suit the mining algorithm. This uses data cleaning, user identification and session identification algorithms.

The data processing identifies the page set in which videos exist for which users search. This set is arranged in time sequence and consists of a home page M, video type F_i and video N_j, built as a search path, as shown in Figure 3. Starting from the homepage of a VOD system, to find a video, two situations can occur:

1. User expectation is consistent with the system classification, and they directly find the target video;
2. Users do not find the target video in the expected category.

If Situation 2 occurs, the other categories are searched. For example, a user wants to find video N_3. The user may look up F_m with no result. Users will return to search other sub-categories for the target video. Finally, the target video will be found from sub-category page F_1, as shown in Figure 3.

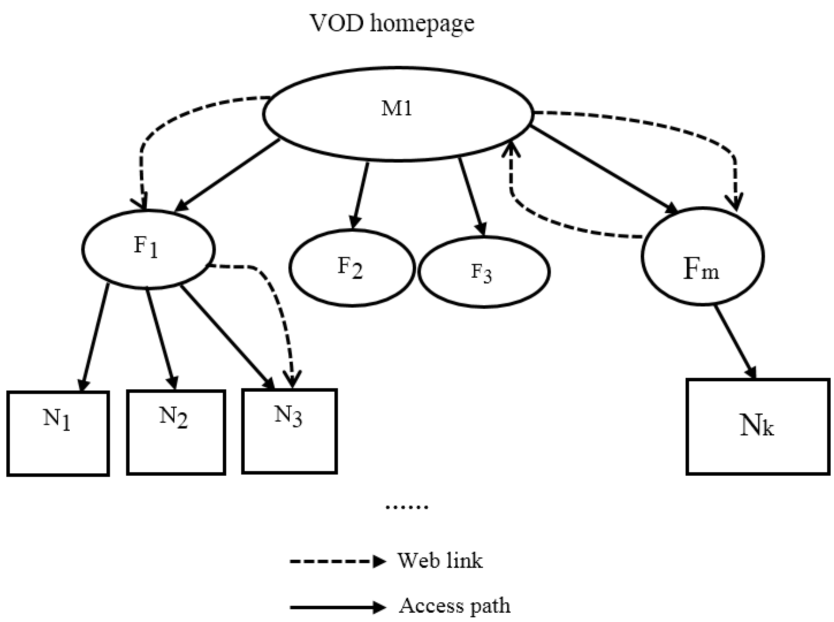

Figure 3: Video search path.

In general, the user always searches for a video using the expected category. If not found, the other classifications are used until the video is found. The user's expectations of the classification of the video depend on the order of the pages used for access. The user access path to the target video provides a reasonable classification of user expectations. On this basis, the video classification can be adjusted to comply with most users' desires.

A video expected category set refers to a default category in the site and a user's access path in the process of looking for the video. A user may access several types when searching for a video, resulting in multiple target video expected categories. Different expected categories based upon Web site hierarchies are determined by user visits to access videos. The expected category set U is calculated from all user access to the video reflecting the different expectations. The set U = {N, (F_{k1}, F_{k2}, F_{k3},..., F_{kn})}, where N is the target video, (F_{kl}, F_{k2}, F_{k3},..., F_{kn}) are classified path sequences when accessing the video. The algorithm is shown in Table 1. One expected video category corresponds to each element in the video expected category set U. To improve the determination of video type, access paths to the video are integrated over a time period. After several experiments, an expectation V was developed to record all users' expected classifications, sorted by each candidate expectation for accessing the video. The formula is as follows:

$$V = (N_k, F_i) = \sum_{j=1}^{x} \frac{X_{ij}}{j} \qquad (1)$$

where X = the total number of categories, X_{ij} = number of times that F_i appears in rank j of all sequences.

In the formula, the weight value $\frac{1}{j}$ embodies the user expectations of the video category. For example, in the expected classification sequence {N_3, (F_m, F_1)}, F_m and F_1 stands for pages visited to find video N_3. For example, a user expects to find N_3 in F_m; hence, j = 1, and $\frac{1}{j}$ = 1. After that, user expectations continue to fall with j constantly increasing, and the weight $\frac{1}{j}$ falling.

User expectation of the video classification is the maximum V of the corresponding video categories. Because of the huge number of videos, many videos need reclassifying after computing, which results in an increased workload to adjust the site. In order to avoid an excess workload, while still reasonably classifying a video, a threshold was set. The video classification can be adjusted, if the ratio between the original classification expectation and maximum expected value is less than the threshold. The provisional threshold value is 0.8.

The algorithm is shown in Table 1.

Table 1: Video classification optimisation algorithm based on Web data mining.

Input	Expected set U generated from videos on demand by users during a time period.
Output	Video sets changed by classification.
Begin	l) Define double dimensional array V in which expected value of each category of all kinds of videos. Attribute m= categories, tuple n=videos, initial value 0;

	2) while(expected category set!= null)do
	read--K(U,N_item,K_item); //read one of the candidate category set;
	// the selected expected category set K_item, the video name N_item;
	power=1; //create the first weight
	for i=0 to K_item.length do
	Power= power+1;
	V [N_item][K_item[i]]=power;
	power=power/2.0; //calculate expected value
	End for i
	End for while

	3) Calculate maximum in V of all videos with formula (1);
	4) If original category set's expected value/maximum<0.8 then output video name and expected category(N_i, F_j)
END	

Experimental Results

The logs on the site were compared before and after they were adjusted using the algorithm. Average times to click a category page when visiting a target video were recorded. The results demonstrate that with the help of the modified algorithm, the average click time was reduced, as shown in Figure 4.

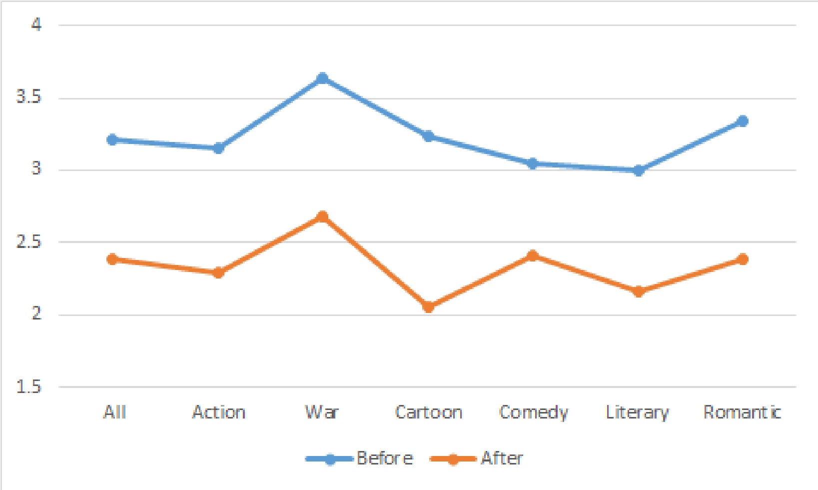

Figure 4: Comparison of average click times.

Experiments demonstrate the advantages of the video optimisation classification algorithm based on Web data mining, with videos totally classified beforehand by VOD system management. Its advantages include improving user participation, better classification, improved visit efficiency and improved service quality of the VOD Web site.

CONCLUSIONS

Video on demand in early development allowed users to select and watch/listen to videos or audio content autonomously for entertainment. Gradually, VOD spread to teaching. Video on demand with base station (BS) framework has been widely used on network platforms at universities and in tele-education systems. With the development of digital television and 4G communications, as well as the blossoming of MOOCs, video on demand has continued to develop. It is possible for users to search by many routes and the on-line system is quick and easy to learn.

These studies investigated the means of extracting information from user dialogue files by Web mining and analysis of the order and rate of category pages that users visited when searching for videos. An optimisation algorithm was developed to better classify videos, so as to reflect user subjective criteria. The VOD site structure can be optimised; thus, reducing VOD site management, saving network costs. Meanwhile, searching efficiency improves. Users are more relaxed checking video on demand and digest the information better than if flooded with learning materials. This facilitates learners' study and development as a result of the intelligent management to on-line videos of learning materials. This study could be a reference guide for, and the development of, MOOCs, as well as increasing VOD influence on tele-open learning and on-line learning by assisting life-long learning in society.

ACKNOWLEDGEMENT

This work was financially supported by the 2015 Educational Reform Research Project of Higher Educational Institutes of Hainan.

REFERENCES

1. Tai, Y., Analysis of college oral English test scores based on data mining technology. *World Trans. on Engng. and Technol. Educ.*, 11, **3**, 180-185 (2013).
2. Han, J., Kamber, M. and Pei, J., *Data Mining Concepts and Techniques*, Peking: China Machine Press, 62-70 (2007).
3. Mao, G., Duan, L. and Wang, S., *Data Dining Theory and Algorithm*, Peking: Tsinghua University Press, 110-112 (2007).
4. Cooley, R., Tan, P.N. and Srivastava, J., WebSIFT: the WEB site information filter system. *Proc. the WEB Usage Analysis and User Profiling Workshop* (1999).
5. Willet, P., Recent trends in hierarchical document clustering: acritical review. *Infor. Processing and Manage.*, 24, 577-591 (1988).