Using data mining technologies to find learning activity rules for online learning

Ke Zhu*

Department of Information Technology, Henan Normal University, East of Construction Road, Xinxiang, China

Received 1 March 2014, www.cmnt.lv

Abstract

Researchers are interested to improve learners' and instructional designers' performance in online learning systems. Learning Activity are of great importance at present, since they are the building blocks of different types of online learning systems. As the number of learning activity grows exponentially and our needs for learning expand equally dramatically, the lack of information or rules about the usability of learning activity places a critical and fundamental constraint on our ability to discover, manage, and use learning activity. This study presents a new approach of data mining and an assessment scheme by combining four computational intelligence theories, i.e., the Clustering Algorithms, the Classification Algorithms, and Association Algorithms, to identify the learning activity rules in online learning systems for learner and instructional designers. Experimental results indicate that the evaluation results of the proposed approach and scheme are improving the work of teachers in designing and searching, and also in the management of Los in a web-based learning environment according to the obtained learning objects usability rules.

Keywords: learning activity, usability rules, data mining, learning path

1 Introduction

As more of our commerce, entertainment, communication, and learning are occurring over the Web, the amount of data generated by online activities is skyrocketing [1]. Commercial entities have led the way in developing techniques for harvesting insights from this mass of data for use in identifying likely consumers of their products, in refining their products to better fit consumer needs, and in tailoring their marketing and user experiences to the preferences of the individual. More recently, researchers and developers of online learning systems have begun to explore analogous techniques for gaining insights from learners' activities online [2].

Around the world, large amounts of money have been spent on the development of courseware, computer applications designed to help people learn. With the advent of the World Wide Web in the late 1990s, it became possible to develop web-based courseware which was more adaptable and cheaper to create, but which lacked the educational and multimedia richness of the earlier monolithic applications [3]. Web is a large-scale service where information is stored in digital format and retrieved over networks and facilitates transmission of the information through internet. In recent years, learning objects as a new information transmission unit has become to be used on web [4].

Learning activities, as the name suggests, are activities designed or deployed by the teacher to bring about, or create the conditions for learning. The difference between a Learning by Design approach to employing various learning activities and other approaches to teaching relates to the pedagogical character or focal intent of the activities selected [5]. What do I want to achieve with this activity? How will I achieve my aims? Which Knowledge Process is best suited to achieving my aim? With Learning by Design the teacher mindfully designs or chooses particular learning activities based on which Knowledge Process is activated by that activity. Some learning activities stimulate experiential learning, others mobilize conceptual thinking, while still others prompt students to engage in analytical discussion. The pedagogical effectiveness of a Learning Element – a teacher's overall design – can be traced to:

1) The mindful selection of learning activities based on the Knowledge Processes which those activities set in motion;

2) The establishment of direct links between those activities and the intended Knowledge Objectives;

3) The careful sequencing of those activities such that they build on, or contribute to, the learning of earlier or later activities Learning Objects (LOs) are small in size, reusable digital resources that can be used with a web browser support. A short and explanatory definition suggested for LOs as interactive web based tools designed to enhance, amplify and guide learning. Polsani proposes an expanded definition about the Los: independent and self- standing learning content units predisposed to reuse in multiple instructional context. Opposed to traditional instructional media which can only be used in one place at a time, LOs can be reused in different contents. Reusability feature is based on the object-oriented paradigm of computer science, which is the main difference between LOs and classic web tools. LOs may be in the form of text, video, audio, graphics or multimedia. Teachers or instructional designers need an environment to manage, store and organize the LOs that are known as Learning

* ***Correspondinf author's*** e-mail: xhukehtu@163.com

Object Repository (LOR). LORs also help teachers or instructional designers create new courses on web that enable the applicability and reusability of (LOR) [6].

For organizations to take full advantage of the potential benefits of learning objects, learning objects must become an integrated part of the instructional technology infrastructure [7]. Managing LOs is generally realized in a structured way as search operations in LORs. In order to create search operations, metadata that means "data about data" is used within LOs. Metadata is information about an object, be it physical or digital. As the number of objects grows exponentially and our needs for learning expand equally dramatically, the lack of information or metadata about objects places a critical and fundamental constraint on our ability to discover, manage, and use objects. Metadata has descriptive in-formation about LOs for finding, managing, using and reusing them more effectively. Perhaps the main distinguishing feature of LO from other tools is their ready availability through web based repositories, collections that can be searched with standardized metadata [8]. Metadata is traditionally found in the card catalogues of libraries. As information has become increasingly digital, metadata is also used to describe digital data using metadata standards specific to a particular discipline. By describing the contents and context of data files, the quality of the original data/files is greatly increased. For example, a webpage may include metadata specifying what language it is written in, what tools were used to create it, and where to go for more on the subject, allowing browsers to automatically improve the experience of users.

Metadata is defined as data providing information about one or more aspects of the data, such as:

- Means of creation of the data,
- Purpose of the data,
- Time and date of creation,
- Creator or author of data,
- Location on a computer network where the data was created,
- Standards used.

For example, a digital image may include metadata that describes how large the picture is, the colour depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

The enormous growth of learning objects on the internet and the availability of preferences of usage by the community of users in the existing learning object repositories (LORs) have opened the possibility of testing the efficiency of data mining technology on recommending learning materials to the users of these communities. It is known that LOR represent a special kind of digital collection, in which the preferences about items can be considered to be related to contingent learning needs. In consequence, seeking learning resources can be hypothesized to be substantially different from selecting products for purchase or information resources. Learning objects have been analysed from numerous perspectives. One area that has been largely overlooked is the learning objects usability rules. It is reasonable that this omission has occurred given that the investigation of learning objects is relatively new. This paper presents the advances in knowledge discovery activity applying the data mining (DM) techniques to the information that can be drawn from LOs.

2 Methodology

2.1 DATA SOURCES IN LEARNING OBJECTS

Information about LOs is diverse and often poorly structured. To conduct studies about usability rules of LOs, it is necessary to have an appropriate framework for studying. In recent years we have focused, among other things, to determine a set of attributes that can rep resent main LOs characteristics as fully as possible. We propose a characterization of LOs, based on four data categories to be processed:

- LOs Metadata
- The internal logical relationship of LOs
- External assessments from questionnaire
- Information obtained from user interacting LOs

2.1.1 LOs Metadata

Learning Object Metadata is a data model, usually encoded in XML, used to describe a learning object and similar digital resources used to support learning. The purpose of learning object metadata is to support the reusability of learning objects, to aid discoverability, and to facilitate their interoperability, usually in the context of online learning management systems (LMS). The IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata is an internationally recognized open standard for the description of "learning objects" [9]. Relevant attributes of learning objects to be described include: type of object; author; owner; terms of distribution; format; and pedagogical attributes, such as teaching or interaction style. This standard specifies the syntax and semantics of Learning Object Metadata, defined as the attributes required to fully/adequately describing a Learning Object. Learning Objects are de-fined here as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. Examples of technology supported learning include computer-based training systems, interacttive learning environments, intelligent computer-aided instruction systems, distance learning systems, and collaborative learning environments. Examples of Learning Objects include multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology supported learning.

2.1.2 The internal logical relationship of LOs

We use thesaurus marking the relationship between Learning Objects. Four relationships are defined as follow: related terms, broader terms and narrower terms. The technical route is first to identify the Learning Objects, describes the relationship between the topics, then uses the theme graph grammar standardization of description, on the basis of the use of the thematic map processor, link resources and to organize these resources, finally to provide Interconnected network relationship map. Figure 1 shows the frame diagram of internal logical relationship of Topic.

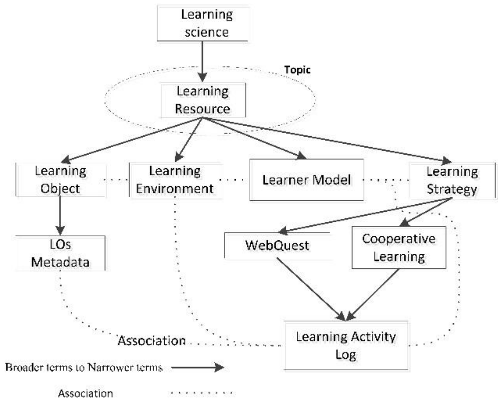

FIGURE 1 The frame diagram of internal logical relationship of topic

2.1.3 External assessments from questionnaire

In September 2012 we began to design a questionnaire that could be used to evaluate the level of inclusion of Learning Object under the support of our secondary development of Sakai learning platform. It was applied to more than 1121 Learning Objects. We consider evaluation around a set of quality indicators of educational multimedia objects, such as the coherence with the learning objectives, quality of the content, feedback, adaptation, motivation, design, presentation, usability, interaction, accessibility, reusability and compliance with standards [10]. The conclusions we obtained from this survey were the basis to redesign the questionnaire and refine the questions to have more and better results of the learning objects.

2.1.4 Information obtained from user interacting LOs

The interactions that Learning Objects have with students, with the instructors, and with educational resources are valuable indicators of the effectiveness of a learning experience. The increasing use of information and communication technology allows these interactions to be recorded so that analytic or mining techniques are used to gain a deeper understanding of the Learning Objects usability rules and propose improvements.

2.2 LO MANAGING SYSTEM

Sakai is a community of academic institutions, commercial organizations and individuals who work together to develop a common Collaboration and Learning Environment (CLE). The Sakai CLE is a free, community source, educational software platform distributed under the Educational Community License (a type of open source license).

This paper proposes a framework of adaptive learning systems which use learning analysis technology. Figure 2 shows the Learning Object Management System Framework. Through the secondary development of Sakai learning platform, we integrated data analysis tool into it to achieve the function of adaptive learning system. The framework aims to provide an infrastructure that supports the development of instructional design activity. Particularly it provides solutions for LOs management.

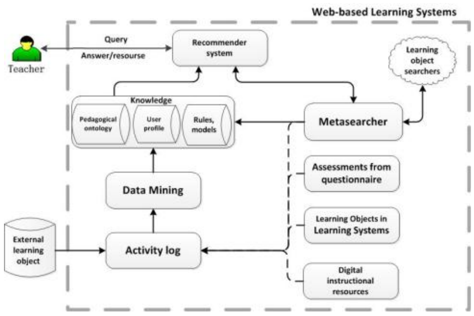

FIGURE 2 The learning object management system framework

2.3 INTEGRATION AND CHECKING DATA

Databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size and their likely origin from multiple, heterogeneous sources [11]. Low-quality data will lead to low-quality mining results. Data integration merges data from multiple sources into a coherent data store such a data warehouse. Careful integration can help reduce and void redundancies and inconsistencies in the resulting set. This can help improve the accuracy and speed of the subsequent data mining process.

2.4 DATA CLEANING

Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. We use a global constant to fill in the missing value. Use a measure of central tendency for the attribute.

2.5 DATA TRANSFORMATION AND DATA DISCRETIZATION

In this reprocessing step, the data are transformed or consolidated so that the resulting mining process may be more efficient, and the patterns found may be easier to understand. Data transformation routines convert the data into appropriate forms for mining. For example, in normalization, attribute are scaled so as to fall within a small range such as 0.0 to 1.0.Data discretization transforms numeric data by mapping values to interval or concept labels. For nominal data, concept hierarchies may be generated based on schema definitions as well as the number of distinct values per attribute.

2.6 APPLY DATA MINING RESULT

The range of possible solutions to the above problems is vast, extending far beyond our approach to instructional technology. While our approach to instructional technology alone cannot resolve all of the problems, it can have a significant impact on all of them. For this stage, we used SPSS（Statistical Product and Service Solutions）that provide data mining algorithms for clustering, classification, and association. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

3 Results

In this section, for each algorithm used in the study, the test characteristic and results obtained are shown.

In each test, the number of clusters was calibrated to generate the greater amount of clusters having mutually exclusive attributes.

Various tests were verified with ID3 algorithms with the already mentioned datasets. We obtained a set of IF-THEN-ELSE rules from the algorithms. After an analysis, we eliminated those rules that were with irrelevant information. Table 1 shows some of the best rules obtained.

TABLE 1 Some of the best rules obtained with the classification algorithms

Rules–Generated	Rules–Interpretation
format =PPT; knowledge transfer =demonstrations; media balance = regular => high	The LO has a high semantic density if it does fulfill the next requirements: has a PPT format contains demonstrations.
format =JPG; media balance = basic; structure =atomic =>low	The LO has a high semantic density if it does fulfill the next requirements: has a JPG format, the media balance is basic.
format =MP3; educational resource type = reading => high	The LO has a high semantic density if it does fulfill the next requirements: has a MP3 format and is used as a reading in class.

The Apriori algorithm is a seminal algorithm for mining frequent item sets for Boolean association rules It explores the level-wise mining Apriori property that all nonempty subsets of a frequent item set must also be frequent At the k-th iteration (for $k \geq 2$), it forms frequent k-item set candidates based on the frequent (k–1), and scans the database once to find the complete set of frequent k-item sets.

TABLE 2 Some of the best rules obtained with the association algorithms

Rules–Generated	Rules–Interpretation
version = final; environment = classroom => aggregation level = 1	If the LO consists of the final version and is used in perennial learning, then aggregation level is basic
structure = atomic; version =final; environment=classroom=>aggregation level = 1	If the LO consists of atomic structure, final version, and is used in perennial learning, then aggregation level is basic
aggregation level= 1; version= final; easy navigation = security => environment = classroom	If the LO is basic and security navigation then it is used in perennial learning

4 Discussion

If all of the properties of a course can be precisely defined in a common format, the con-tent can be serialized into a standard format such as XML and loaded into other systems. When it is considered that some e-learning courses need to include video, mathematical equations using MathML, chemistry equations using CML and other complex structures the issues become very complex, especially if the systems need to understand and validate each structure and then place it correctly in a database.

This study investigated the rules which using Learning Objects. The findings generally derived from the model as well as earlier empirical studies. As mentioned earlier, we use LOs Metadata, the internal logical relationship of Los, external assessments from questionnaire and information obtained from user interacting Los as studying data source. It is possible for example, to package learning objects with SCORM specification and load it in adaptive learning systems which use learning analysis technology.

The findings indicate that the use of methods and data mining techniques are useful for the discovery of knowledge from information available in LOs. The result allows us to establish what elements are crucial to classify suggest, or recommend action values in a learning management system. Teachers are aware that they must plan learning experiences that match the differing competencies of their students. The findings also show that many teachers, however, are not familiar with how learning objects can make demands on students' literacy, memory, and cognitive abilities. A mismatch in any area will result in a less successful learning experience for the student. Teachers need to understand this for both selection and use of learning objects.

5 Conclusions and Outlook

In this paper, we presented an adapted methodology for the application of data mining techniques to Los, trying to discover relevant rule in its design and usage characteristics. This can help teachers to assess learning resource precisely utilizing learning objects usability rules in a web-based learning environment. Additionally, teachers can devote themselves to teaching and designing courseware since they save a lot of time in evaluating learning objects. More significantly, teachers could understand the factors influencing learning performance in a web-based learning environment according to the obtained interpretable learning objects usability rules. Students will have the benefit of access to the highest-quality learning resources avialable, making a significant impact on the quality of their learning experience and their learning outcomes.

Further collect data from other channel are required in future work. For example, it would be interesting to consider the situational of learning as an element to recommend teachers and students suitable learning objects with real environment. Also, it would be consider the location, equipment, even weather and something else related with learning.

Acknowledgments

This work is supported by soft-scientific project of Henan Province (No. 142400410425).

References

[1] Boyle T 2009 The handbook of research on learning design and learning objects Wollongong: IGI Global Snippet **2** 311-5

[2] Brosnan K 2005 Developing and sustaining a national learning-object sharing network: A social capital theory perspective *Proceedings of the ASCILITE 2005 Conference* 105-14

[3] Chitwood K, May C, Bunnow D, Langan T 2000 Battle stories from the field: Wisconsin online resource center learning objects project: http://reusability.org/read/chapters/chitwood.doc

[4] Kanellopoulos D, Kotsiantis S 2012 Evaluating and recommending Greek newspapers' websites using clustering *Program: Electronic Library and Information Systems* **46**(1) 71-91

[5] Zhu K, Liu Q 2012 The Acceptance of Sakai Technology by the Chinese College Students *Advances in Information Sciences and Service Sciences* **4**(17) 435-41

[6] Menendez-Dominguez V, Zapata A, Prieto-Mendez M E, Romero C, Serrano-Guerrero J 2011 A similarity-based approach to enhance learning objects management systems *11th International conference on intelligent systems design and applications* Cordoba 996-1001

[7] Friesen N 2005 Interoperability and learning objects: an overview of e-learning standardization Inter disciplinary *Journal of Knowledge and Learning Objects* 23-31

[8] Polsani P R 2013 Use and abuse of reusable learning objects *Journal of Digital Information* **3**(4) 164-6

[9] South J B, Monson D W 2012 A university-wide system for creating, capturing, and delivering learning objects Wiley D A (Ed) *The Instructional Use of Learning Objects Online Version. Retrieved: http://reusability.org/read/chapters/south.doc*

[10] Normey Rico J E, Camacho E F 2009 Unified approach for robust dead-time compensator design *Journal of Process Control* **19**(1) 38-47

[11] Cao Y, Zhang G, Su Y 2004 Mathematical modelling of helicopter aerobatic manoeuvres *Aircraft Engineering and Aerospace Technology* **76**(2) 170-8

Author

Ke Zhu, 1982, Henan Province, China

Current position, grades: lecturer.
University studies: master's degree in major of educational technology, Henan Normal University in 2008.
Scientific interests: neural network algorithm, data mining.