Personalised E-Learning Opportunities – Call for a Pedagogical Domain Knowledge Model

Alex G. Büchner
Synergy Learning
1 Millennium Way
Springvale Business Park
Belfast BT12 7AL, UK
alex.buchner @ synergy-learning.com

David Patterson
University of Ulster
Northern Ireland Knowledge Engineering Laboratory
Faculty of Engineering
Newtownabbey, BT37 0QB, UK
wd.patterson @ ulster.ac.uk

Abstract

A considerable amount of e-learning content is being delivered via virtual or managed learning environments. These platforms keep track of learners' activities including content viewed, time spent and quiz results. This monitoring trawl provides appropriate data to enable personalised e-learning experiences through the application of existing data mining and knowledge discovery techniques.

The reasons for not providing this type of bespoke teaching is, in addition to technical and financial constraints, largely due to the plethora of educational and pedagogical issues which have to be overcome. This paper presents these obstacles and suggests solutions thus challenging the community to focus on a new research area which concentrates on facilitating the specification and application of pedagogical domain knowledge for incorporation into existing data mining and knowledge discovery frameworks. This includes educational thresholds, constraints, taxonomies and previously discovered knowledge as well as pedagogical interestingness and metrics.

1. Introduction

An established research community has emerged over the last decade from the synergistic fields of education and artificial intelligence. This has provided a valuable body of research which has been disseminated through such forums as the International Journal of Artificial Intelligence in Education and the International Conference on Artificial Intelligence in Education. Recommendation technologies, profiling and related data mining and knowledge discovery applications are common contributions from within this research community.

At the same time, the range of e-learning implementations has increased substantially both in academia and industry. The majority of educational content is being delivered via virtual or managed learning environments (VLEs), such as WebCT, Blackboard and Microsoft Class Server. In addition to the standard facilities such as content management, course organisation, collaboration and assessment, all systems provide some degree of logging to monitor the progress of learners including content viewed, time spent at a particular subject and quiz results. This monitoring trawl provides ideal data to personalise e-learning experiences by applying data mining and knowledge discovery techniques.

Given these circumstances, one would expect that learners would enjoy a fully personalised learning experience, whether they have signed up for a commercial skills course or enrolled in a formal qualification at a further or higher education institution. The reality however, is that learners are often presented with the same learning style that learners have been presented with for 50 years. An instructor – in whatever form – designs course material which has to be followed in a pre-assigned sequence. The fact that much of the material is now multi-media based and as such can be consumed at the student's own pace giving them more autonomy and control, has had little impact on the style of learning encouraged, which is still instruction led (linear) as opposed to knowledge led (non linear).

The objective of this paper is to show the potential of applying data mining and knowledge discovery techniques to e-learning and to establish the reasons for the reluctance of some educators to introduce recommendation technologies which would allow the move from instructive online teaching to constructive e-learning. In order to overcome this discrepancy, a go-between is proposed in the form of pedagogical domain knowledge which allows the specification of the instructor's experience and targets to be achieved to be taken into account by the recommendation mechanisms. This can be expressed in the form of educational thresholds, constraints, taxonomies and previously discovered knowledge.

2. Related Work

2.1. Knowledge Engineering Approaches

Substantial academic resources have been investigated in the modelling of learner's behaviour in different environments and contexts. A representative example is the Global Campus project [1]. While this information is useful from a teaching perspective, it often provides intangible information for personalised learning.

Adaptive hypertext systems have attracted some interest in that they provide a mechanism to dynamically offer links to information that is of interest to users. Interesting work has been presented by [2], in which machine learning algorithms (mainly clustering) have been applied to a learner's previous activities and context-sensitive links proposed.

While substantial research has been carried out in the field of applying data mining algorithms to web usage data in the form of log files [3, 4], relatively little attention has been drawn to its application to data in online learning environments. [5] have proposed the application of statistical techniques and the detection of associations to discover the effectiveness of web-based learning.

[6] have applied collaborative filtering to clusters of learners for the recommendation of what papers to read. The system is unique in that it includes external web sources, it contains a base of curriculum knowledge, and it takes into account pedagogical metrics, such as frequency and paper ratings.

The field of curriculum sequencing has also been studied. High-level or knowledge sequencing determines the next concept or topic to be taught, while low-level or task sequencing determines the next learning task to be delivered [7]. A wide range of intelligent solutions has been proposed for both types of sequencing by a number of researchers.

2.2. Virtual Learning Environments Reports

There are three key competitors in the VLE segment, namely WebCT, Blackboard and Microsoft's Class Server. Additionally, a range of smaller, often local, competitors exist. What all products have in common is that they offer a degree of core functionality (content management, course organisation, collaboration and assessment) and most claim to be SCORM compliant. Additionally, all products offer some form of reporting, which is of interest in this context. There are two types of reports that can be produced:

Server-based statistics utilises VLE or web server log files and runs analysis modules over them. While these statistics are useful, they are only of limited pedagogical value. Examples of statistics that can be produced are

- Number of learners in a day / week / month
- Peak-times of usage
- Top *n* pages / hits / subjects
- Error reports

Leaner-based statistics offer more learner-centric analyses. Similar to its server-based counterpart, logging information forms the data basis, but metrics that are tied in with module structures are used to produce information. Example statistics are

- Time students have spent on a module or unit
- Results of questions that have been answered
- Number of learning objects that were used
- Progress of students in learning entities
- Recency & Frequency information

Data warehousing like operations, such as drilling-down, dicing and slicing, allow the user to tailor analyses to their needs. This type of information – while basic in nature – provides more pedagogical feedback, but lacks the sophistication that is required for appropriate recommendation and personalisation.

3. Data Mining and E-Learning

There exists a wide range of ideas as to how e-learning experiences can be improved by the utilisation of appropriate data mining techniques. Here only a few representative applications and techniques are presented, which demonstrate different forms of individually tailored learning.

3.1. Data Mining Techniques

By applying *classification* algorithms, it is possible to distinguish between learners that are at risk of failing a module and those that are not. This type of knowledge is useful when an instructor has to identify all learners that are underperforming. For example,

```
IF (biology_mark < 40%) AND (chemistry_progress < 25%)
THEN risk = HIGH
(SUPPORT = 42.2%, CONFIDENCE = 26.1%)
```

Using *segmentation*, clusters of learners can be identified, which have similar characteristics. Groups with similar features can be offered specially tailored content to catch up in weak areas or further improve strong areas. A sample labelled cluster can be described as follows:

Strong vocational, weak academic:
(Photoshop > 65%) AND (Flash > 60%) AND (Media Studies < 45%) AND (Final Major Project = Finished) AND (Age < 23)
(WEIGHT = 0.17, SIMILARITY = 0.42)

Associations provide information about learning objects that have been studied in conjunction (as in basket analysis). The most interesting associations are the ones that contain items from different modules or subjects, given that they were not presented together, for instance

Textures, Lights, Geometry, 3D scenes → Marketing
(SUPPORT = 2.2%, CONFIDENCE = 28.6%)

Sequences show in which order materials have been studied. Again, the most interesting sequences are the ones where the student deviates from the presented learning pathway, for example

module1_1 → module1_2 → module1_3 → module4_2 → module4_3 → module1_4
(SUPPORT = 8.5%, CONFIDENCE = 22%)

3.2. Profiling and Recommendation

Based on knowledge created by data mining techniques as described above or specialist mechanisms such as collaborative filtering, it is possible to create learner profiles and provide individual recommendations.

Profiles describe groups of learners with similar characteristics. Example groups are 'novice learners', 'high achievers', 'strong vocational', and so on. Each profile is described by a set of variables and allotted values or value ranges. Profiles can be used for describing a learner's characteristics or recommending learning content.

Recommendations are individually created suggestions of content, which can be in the form of dynamically created links or content. In its most advanced form, this is close to a one-to-one tuition scenario where an instructor provides individually tailored teaching. Recommendation engines usually deduct their actions from a knowledge base, which has been built over time.

3.3. Knowledge Bases

A knowledge base provides a pool of information within a certain domain, where the knowledge is represented as patterns, rules, profiles or any other form of knowledge. In the context of education, knowledge can be acquired based on a learner's previous behaviour or other 'similar' learner's patterns and feedback that has been provided to recommendations made by the system. Additionally, pedagogical experts can incorporate existing knowledge, rules, constraints and so on.

A properly constructed and well maintained knowledge base in an educational setting can provide the ideal companion for any instructor, but also for any (semi-) automated learning system.

4. Reluctance by Pedagogy

Understandably, there exists some reluctance throughout the pedagogical profession when artificial intelligence mechanisms are discussed. As with any other profession that involves a high degree of expertise, the introduction of computer-based methods is being approached with suspicion and teaching is no different. This section outlines the key issues which are raised by the teaching profession, covering off-line, mixed mode and online delivery of academic, vocational and occupational subjects.

The *perception* of some instructors (moderators, teachers or lecturers) is that artificial intelligence based techniques have the potential of replacing some or all of their core competencies. While certain topics can be learned in an instructor-free environment, the majority of content will be delivered in a mixed-mode, moderator-supported or face-to-face teaching style.

There are considerable *pedagogical* issues. Some elements have to be learned sequentially (for instance, in programming: variables, increments, loop), while others can be picked up in any order (for loop, while loop, until loop). In certain situations learners should be confronted with different learning styles (listening, watching, reading, etc). This further requires the instructor to develop multiple pieces of content to provide such variety of learning.

A serious range of *legal* issues has been debated. Imagine two students who, based on their previous learning behaviour, have been recommended different materials during the coverage of a certain subject. One passes the exam successfully, the other doesn't. What are the rights of the failed student? In a course where the learner is a paying customer, can it be expected that everybody is being presented with the same content? This leads to issues around *quality assurance*. How can it be monitored, how can it be transparent and how can it be verified? Currently there exist no conclusive answers to those concerns.

The authors propose that, most issues which cause reluctance by pedagogical professionals are not caused by a phobia of technology, but by the perceived incompetence of current personalisation technology. Whilst not every issue can be resolved by the approach presented in this paper, we believe that some aspects of the shortcomings identified can be overcome. The key to this is through better utilisation and through modelling of pedagogical domain knowledge.

5. Pedagogical Domain Knowledge

Currently, educators provide the content they feel is most appropriate for their learners in a VLE. Independently, knowledge engineers are advocating processes which take into account the log files of the VLE and provide recommended content based on the learner's data and some pre-existing knowledge. These two disparate modi operandi are depicted in Figure 1 below.

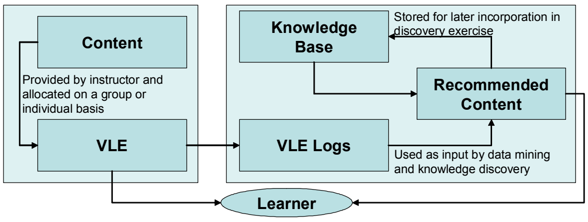

Figure 1. Current Educational Content Provision

It is obvious that there exists a considerable gap between the way learning content is being offered and the research that has been conducted. The key discrepancy seems to be the lack of integration of required achievements and pedagogical knowledge into analysis and recommendation processes.

With the introduction of domain knowledge a range of benefits can be achieved. Technically, they reduce the search space and enable generate more useful knowledge (higher quality, lower quantity) in a shorter discovery period. Pedagogically, it allows the incorporation of expertise and educational limitations, and improves interpretability and simplicity of results.

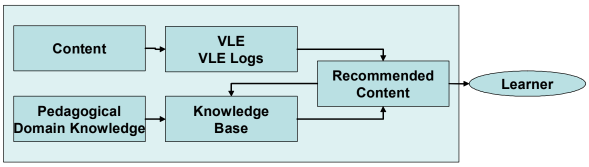

Figure 2. Domain Knowledge driven Content Recommendation

Figure 2 outlines a high-level schema which provides a solid basis for the specification and integration of pedagogical domain knowledge, which can be provided in different formats. A commonly accepted organisation of domain knowledge is that used by [8], which distinguishes between thresholds, constraints, taxonomies and existing knowledge. This structure is used in the following sub-sections.

5.1. Educational Thresholds

In a data mining, context thresholds provide user parameters which allow the introduction of bias to discovery.

Typical examples of algorithm-dependent thresholds are minima and maxima of support, confidence and coverage. Whist almost all thresholds are domain-agnostic, their settings are context-sensitive.

In the context of education, these have to be adjusted according to general pedagogical goals, system identified specific user learning needs (personalisation of content) and delivered subject or legal implications. For instance, a "high support – medium confidence" threshold combination would be used when identifying mainstream learning patterns, while a "low support – high confidence" combination would be selected when identifying positive and negative outliers.

5.2. Educational Constraints

Constraints are restricting conditions which exclude unreasonable or useless information from the discovery and restrict large hypotheses spaces. Syntactic constraints, which have the objective of restricting the search effort, are usually specified using thresholds (see above). Domain constraints have the objective to restrict the number of results. Examples in the education domain are the mapping of content to learning outcomes or time scales in which lessons have to be completed. Quality constraints have the objective of excluding uninteresting results. This can either be achieved through thresholds and interestingness measures or by providing specific pedagogical criteria, for instance, the exclusion of learning activities during the induction period.

The specification of educational constraints depends highly on the type of analysis that is carried out, the learners' data being analysed and the bias that has to be incorporated. In order to allow the comparison of results, it has proven useful to apply similar constraints across discovery exercises.

5.3. Educational Taxonomies

Taxonomies can be described as systems of connected classes where each class contains similar items. The most popular taxonomy is hierarchical, where all classes are organised in a tree. However, other graph structures, such as networks and rings have also proven useful.

In an educational context there is a range of taxonomies which can be built and used as domain knowledge for knowledge discovery. Examples include grouping of units to a subject, course structures, class organisation or content topology in a VLE.

5.4. Educational Knowledge

The objective of incorporating previously discovered or existing educational background knowledge to support the discovery process is to increase the likelihood of finding unexpected information, to avoid contradicting knowledge and to increase the actionability of results.

The formats of educational knowledge that can be expressed depend on the type of knowledge that is being discovered (see Section 3). The types of knowledge that can be incorporated is virtually endless and depends on the context of the search, the objective of the analysis and the bias and experience of the experts involved.

5.5. Future Work

It is obvious from the description of different types of domain knowledge that a host of research and development is required to achieve an acceptable level of personalised e-learning experience. It is compulsory that these endeavours are undertaken in a *synergistic* way, that is knowledge engineers and pedagogical experts have to work in tandem.

It is further imperative that any specifications are product and vendor independent. *XML* (in conjunction with XML schemata and DTDs) has proven the most appropriate language [9]. It (pedXML?) allows the flexible and extensible specification of not only domain knowledge but also of other pieces of information that have to be modelled, for instance courses, lecturers, students, marking schemes, assessments, and so on.

Two closely related subjects that should be studied in the context of education and knowledge discovery are interestingness and metrics. Are existing subjective and objective *interestingness* measures sufficient when modelling pedagogical scenarios? If not, what are amendments and alternatives? What are useful educational *metrics*? Web marketers have developed more than 100 e-metrics in recent years describing phenomena such as slipperiness, seducible moments or revolving door ratios. Examples of potential edumetrics(?) include traditional information such as class average or failure rate and also new information such as online content coverage, page impression to mark ratio or course fee to success rate quotient.

6. Conclusions

Whist proven data mining technology is available to provide 'limited' personalised commercial and academic e-learning opportunities, it is apparent that a range of pedagogical issues have to be resolved in order for it to truly deliver on its expectations. In order to support this process it is necessary that a mechanism is provided that allows the modelling and integration of pedagogical domain knowledge. This gap has to be filled by research from both the knowledge engineering and pedagogy camps. The objective should not be to apply technology for the sake of its application, nor to deliver it by the preferred teaching method and style of the instructor; the goal must be a more flexible, efficient, contextualised and adept learning environment. In a time in which learners have become paying customers, they can expect to be treated equally to their counterparts in retail, financial services or telecommunications. Personalised e-learning will offer such a state-of-the-art experience.

Acknowledgements

The authors would like to thank Rodney McMaster and Alan Largey for their valuable comments.

References

[1] M. Dimitrova, C. Sadler, S. Hatzipanagos, A. Murphy, "Addressing Learner Diversity by Promoting Flexibility in e-Learning Environments", *Proc. 14th Int'l Workshop on Database and Expert Systems Applications*, pp. 287-291, 2003.

[2] A.S.G. Smith, A. Blandford, "ML Tutor: An Application of Machine Learning Algorithms for an Adaptive Web-based Information System", *Int'l Journal of Artificial Intelligence in Education*, Vol. 13, 2003.

[3] M.D. Mulvenna, S.S. Anand, A.G. Büchner, "Personalization on the Net using Web Mining", *Communications of the ACM* (guest editorial for special issue on web mining), 43(8):30-34, 2000.

[4] R. Kohavi, F. Provost. (eds) "Applications of Data Mining to Electronic Commerce", Kluwer Academic Publishers, 2001.

[5] I.S.Y. Kwan Lau, J. Fong, "Investigation on the Effectiveness on Web-based Learning Using Web-Mining Approach", *Proc. 14th Int'l Workshop on Database and Expert Systems Applications*, pp. 302-316, 2003.

[6] T.Y. Tang, G. McCalla, "Smart Recommendation for an Evolving E-Learning System", Proc. 11th Int'l Conf. on Artificial Intelligence in Education (AIED'2003), pp. 699-710, 2003

[7] P.L. Brusilovsky, "A framework for intelligent knowledge sequencing and task sequencing", Proc. 2nd Int'l Conf. on Intelligent Tutoring Systems, pp. 499-506, 1992.

[8] W. Klösgen, J. Zytkow (eds.), "*Handbook of Data Mining and Knowledge Discovery*", Oxford University Press, Chapter 18, 2002.

[9] A.G. Büchner, M. Baumgarten, M.D. Mulvenna, R. Böhm, S.S. Anand, "Data Mining and XML: Current and Future Issues", Proc. 1st Intl'l Conf. on Web Information Systems Engineering, pp127-131, 2000.