Data Mining Technology for the Evaluation of Learning Content Interaction

Abstract. Interactivity is central for the success of learning. In e-learning and other educational multimedia environments, the evaluation of interaction and behaviour is particularly crucial. Data mining – a non-intrusive, objective analysis technology – shall be proposed as the central evaluation technology for the analysis of the usage of computer-based educational environments and in particular of the interaction with educational content. Basic mining techniques are reviewed and their application in a Web-based third-level course environment is illustrated. Analytic models capturing interaction aspects from the application domain (learning) and the software infrastructure (interactive multimedia) are required for the meaningful interpretation of mining results.

Keywords: e-learning, educational multimedia, evaluation, data and Web mining, interactivity, content interaction.

1 Motivation

Interactivity is a central concept in educational environments (Sims, 1997). It refers to the interaction of a learner with the learning material, the instructor, or with peers in the process of learning (Moore, 1992). Interaction in different forms is known to be beneficial for the learning experience and the overall effectiveness of learning. However, a clear definition is still not unanimously agreed upon, which makes the instructional design for and the evaluation of interactive teaching and learning environments issues of ongoing concern.

A fundamental question in relation to instructional design in the context of e-learning and other computer-supported teaching and learning is how learners interact with educational multimedia; that is, what their concrete behaviour, what their preferred learning style, and what their learning goal in such an educational environment is. Answers to these questions are important for integrated formative evaluation and instructional design (Hirami, 2002). The questions are, however, more difficult to answer, if, as in the case of computer-based teaching and learning, direct contact between learner and instructor or between learners is reduced and less feedback is available for the instructor (Northrup, 2001). Another difficulty is created by the potential of the Web and other educational multimedia to enable novel and innovative forms of teaching and learning that are, consequently, not always well understood (Ohl, 2001).

Data mining (Chang, Healy, McHugh, & Wang, 2001) shall be proposed as the central, observation-oriented evaluation technique, which can provide an important contribution to the understanding of learner interaction and instructional design. Data mining is used in a variety of domains from business-oriented decision support systems to scientific data analysis. Data mining is a technique that allows the discovery and extraction of latent knowledge, such as learners' behavioural patterns and usage rules in interactive educational systems, from a computer system's access logs. With data mining, essential activities can be captured, learner behaviour determined, and this behaviour interpreted in the context of learning styles and goals. The advantage of data mining over classical evaluation techniques such as surveys and observations is that it is an objective, non-intrusive technique that allows constant monitoring and evaluation at any time.

The ubiquity of Web and Internet technologies, which are now the predominant technologies for computer-based teaching and learning (Weston & Barker, 2001), has resulted in a special form of data mining called Web mining – the use of data mining in Web and Internet system analysis and evaluation. Web mining has been used extensively in e-commerce applications to analyse customer behaviour. It has proved a powerful tool in improving Web site structures, managing customer relationships, and in providing personalised features – achieving better shopping experience for customers and improved sales for companies. Web mining, however, has, with a few exceptions (Zaïane, 2002; Pahl & Donnellan, 2002), not been deployed in e-learning, despite the potential of being equally valuable in improving the learning experience for learners and in enabling an improved learning process for instructors. Content interaction – central in Web-based teaching and learning – shall be the focus.

The key objective here is to introduce data mining technology and demonstrate its potential for e-learning and other computer-supported teaching and learning systems. E-learning differs from for instance e-commerce in that learning as the goal is more complex and, based on cognitive processes, more difficult to capture than shopping and sales. Interactivity is the central concept. Therefore, educational models of interactivity and learning, but also the technical aspects of interaction of learners with educational multimedia software are central. How these different approaches to interactivity can be integrated and applied as a data mining-supported analytic model for the evaluation of learning shall be illustrated. In order to benefit from the full potential of data and Web mining, we need to be clear about what exactly the evaluation goal is and what types of learning-related information we can obtain for an analysis. Principles of learning and interactivity, in particular in computer-based educational environments, need to be understood and made explicit. Investigating interactivity in all its aspects is a necessary prerequisite for the successful deployment of interactivity mining techniques.

2 Principles of Data and Web Mining

Data mining is concerned with the discovery and extraction of latent knowledge from a database (Chang, Healy, McHugh, & Wang, 2001). Typically, this knowledge is classified into rules and patterns that can help an analyst in analysis and decision making processes. Data mining has been used for a wide range of applications ranging from decision support systems in business applications to analysis tools in scientific applications. The purpose of data mining can be predictory (decision support), generative (create new/improved designs), or explanatory (scientific analysis).

Web (usage) mining is the analysis of (user behaviour) data in Web-based systems. The database is the access log created by a Web server. The fact that only activities are recorded makes Web usage mining different from data mining in general. Each Web request of any text document or other type of resource is recorded in the access log. Web log entries reflect activities – the requestor, access type, access time, and requested resource are recorded. In education-specific terms the learner, form of activity/interaction, access time, and content item are recorded.

A range of classical data mining techniques exist that support the extraction of rules and patterns from a database (Agrawal & Srikant, 1995; Cooley, Tan, & Srivastava, 1999; Chang, Healy, McHugh, & Wang, 2001):

- *Usage statistics* are usually not considered as data mining techniques. However, they often form the starting point of evaluations. For Web-based systems, usage is captured in simple statistical measures such as total number of visits, number of visits per page, and so forth. Tracking features of most e-learning platforms are based on these measures.
- *Classification* and *prediction* are related techniques. Classification predicts class labels, whereas prediction predicts continuous-valued functions. A model is used to analyse a sample. The result of this learning step is then applied. Regression is a typical form of prediction.
- *Clustering* groups mutually similar data items. In contrast to classification, the class labels are not pre-given. The learning process is called unsupervised in this technique. Pattern recognition is a typical example.
- *Association rules* are interesting relationships discovered among the set of data items. A typical example is purchasing analysis, which can identify item pairs frequently purchased together.
- *Sequential pattern* analysis is applied if events are captured in a database over a period of time. Frequently occurring patterns are extracted. Web usage or sales transaction patterns are typical examples.
- *Time series*, the analysis of the variance of patterns and rules over time, are important since they allow the analyst to evaluate changing and varying behaviour.

Often, a session, which is a period of uninterrupted usage, is the basic unit of analysis.

The understanding of levels of abstraction of knowledge and of languages to express knowledge is critical for the success of the data mining technology. Concepts, e.g. learning activities and interactions, have to be clarified. A Web log contains low level concept information. Moving up the concept hierarchy of learning interaction means to move from a technical reflection of interaction to learning activities, tasks and goals. Summarising and generalising individual behaviour into behavioural patterns and the interpretation these patterns in a domain-specific analytic model is essential for an evaluation.

In the implementation of the mining techniques, different phases can be distinguished – see Figure 1.

- After a *cleaning* phase, in which irrelevant entries such as images in text pages are removed, *rules and pattern extraction* is the first core phase based on mining techniques introduced earlier on. Abstraction and retrieval functions are performed. Often the first selection of patterns and rules has to be reduced to a tractable size by threshold control, identifying meaningful and interesting patterns and rules only.

- *Interpretation* is the second core phase where patterns and rules have to be interpreted in an *analytic model*. Only in the context of the application domain it can be decided what a meaningful mining result is and how these results have to be interpreted. The interpretation is therefore guided by an explicit or implicit analysis model. If the focus is on usage and user behaviour as in Web-based systems, *languages* are important in these analytic models. For instance, Web log and interactivity languages allow us to formulate rules and patterns of user interaction.

A few limitations apply to Web mining. A classical problem is caching. Web browsers keep copies of already visited pages in a cache. Further requests of these pages will be satisfied from the cache and do not produce entries in the Web server's access log. However, techniques to deduce the correct patterns exist (Zaïane, 2001). Furthermore, most advanced systems use dynamically created pages, which circumvents this problem. Some recent technologies such as mobile agents require changes to the way access logs are created since computation is moved from the server to the client. Web mining is limited to activities occurring under control by the Web server. Activities beyond that site are not captured and cannot be mined.

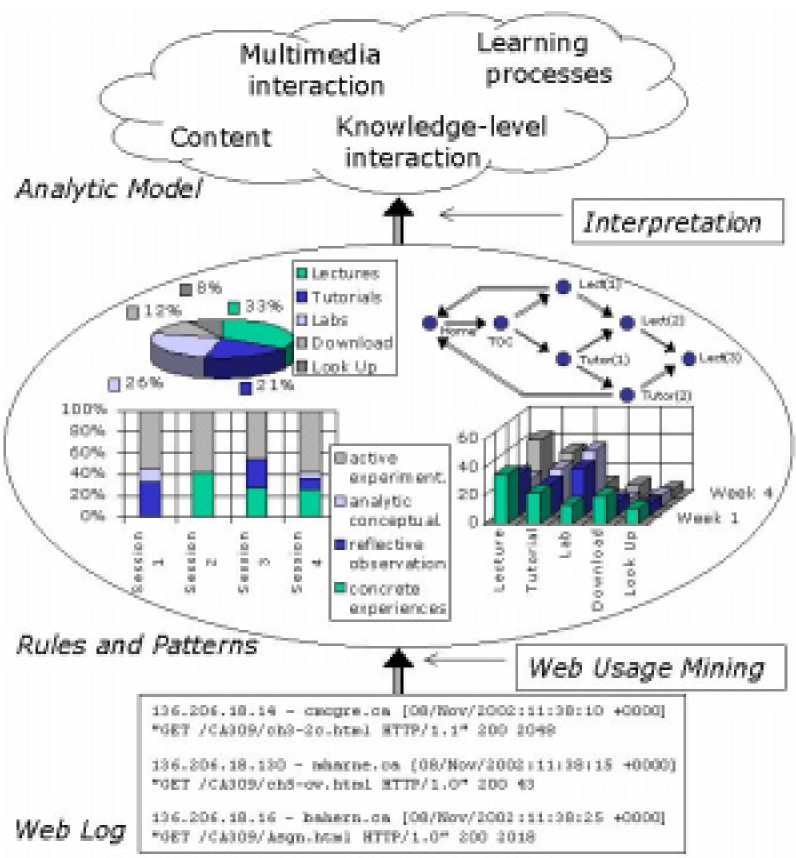

Figure 1. Web Usage Mining and Interpretation Architecture.

3 Interaction

Interaction is here the central concept – a concept that has meaning on different layers ranging from the learning domain to the technical infrastructure. Three perspectives on interaction shall be offered: *(i)* interaction and learning, *(ii)* human-computer interaction, and *(iii)* interactive educational multimedia.

3.1 Learning and Interaction

It has long been argued that learning should be an active process (Dewey, 1916). Interactivity is an essential ingredient in this process that positively affects learner satisfaction and performance. Even though various authors have addressed this issue, an accepted definition of interaction for the various educational environments, in particular addressing educational software, is still an open question. A model of the interactions that take place between a learner and what the learner is trying to learn can help instructional designers provide the learner with a diverse set of interactions that foster learning. Such a model also supports the instructor in the evaluation of interactive teaching and learning. Models provide the foundations for an approach to instructional design; they clarify what the learning activities and what the knowledge representations are.

Some authors organise interaction by roles (Moore, 1992), (Norman, 1998). Moore introduces three types – learner-content, learner-instructor, and learner-learner interactions. Some attempts have been made to revise the role-based models in favour of a more content- or knowledge-based definition (Jonassen, 1994; Sims, 1997). The interaction with content, which represents subject knowledge, has a central function, in particular if we move away from the classical classroom-based educational environment with teacher and peer interaction towards computer-based education (Ohl, 2001). Interaction is an internal dialogue of reflective thought that occurs between learner and the material. Interaction is triggered and supported by events in the learning environment, such as interaction with computer-based educational media. Ohl (2001) focuses on how the learner interacts with what is to be learned; these are often the only two core elements that remain if we leave the classical educational environment.

3.2 Human-Computer Interaction

The definition of learning as a dialogue between learner and content needs to be adapted to the human-computer environment. Norman's (1988) model of interaction between human and computer introduces an *execution-evaluation cycle*. The user formulates a plan, which is then executed by the system. The user observes the interface to evaluate the results and to determine further actions. We use models to capture principles of interaction. A model's purpose can be both evaluative (support in evaluation) and generative (support in design). *Cognitive models* represent the knowledge, intentions, and processing abilities of users (Dix, Finlay, Abowd, & Beale, 1993). In particular, the acquisition and formulation of plans of activity through a hierarchical model of goals and tasks and the execution of plans of activity through a linguistic model need to be addressed.

A *hierarchical model* captures the user's task and goal structure. A *task* is an operation to manipulate concepts of the domain. A *goal* is the desired output from a task – to be accomplished in some domain. A goal and task hierarchy is defined by dividing goals into subgoals and tasks. A user accomplishes goals by solving subgoals. *Strategies* define how goals on the same level are connected and scheduled. Strategies are the expression of learning styles and study habits. The acquisition and formulation of goals and tasks is based on *knowledge-level* attributes relevant to the user. This will then be mapped into actions and attributes relevant to the system. Data mining will turn out to be ideally suited for the discovery and extraction of goals, task execution, and interaction behaviour.

A *linguistic model* focuses on the user-system grammar constraining the interaction. This takes into account the different *styles of interaction* – direct manipulation, command line, or form fill – that characterise the form of manipulation of concepts represented by interface elements. Languages that capture the process of interaction serve different purposes, for example to analyse the cognitive difficulty of the language or to describe dialogues. A dialogue specification, which captures the recurring cycles of execution and evaluation, defines what the legal user actions and system responses are. The language is central since user activities and dialogues can be extracted from Web logs through data mining.

Cognitive architectures help us to address cognitive learning processes in the context of human-computer interaction. The architecture that a system provides to allow a user to accomplish a goal defines a problem space, the actions that can be used to traverse this space, and a set of desirable states that represent the goal. The problem is the subject; the goal is to learn about the subject. Educational systems incorporate paths to successful learning; they characterise possible solutions to the learning problem in form of structure and dialogues. The user provides a solution how to traverse to the desired state. We can distinguish two levels of interaction and activity: *knowledge-level activities*, which are tasks and actions based on the user's perception of the system, and *problem-space processes*, which involve goal formulation, action selection, action application, goal completion. Knowledge-level goals can be inferred from the formulation of problem space-level goals, for example through Web mining. Behaviour at the knowledge-level can be inferred from the operation selection.

3.3 Interactive Educational Multimedia

In computer-based education, the interaction of a learner with what is to be learned, presented by the educational software, is central. Usually, educational software provides the learner with various forms of communication, often using different individual media. Therefore, we use the term *interactive educational multimedia* to emphasise the interactive nature of educational software and the variety of activities and interactions. Understanding educational software as *interactive multimedia* systems is a key to understanding learning and interactivity in computer-based educational environments. Multimedia software is designed for interaction with the user and with usability in mind (Elsom-Cook, 2001). Principles of multimedia systems explain issues related to knowledge that is represented, activities that are offered, and types of interactions that are possible.

Multimedia systems are characterised by the communication *channels* that are provided for users to access and communicate knowledge. The user uses specific *languages* to communicate along these channels. The interface that allows a (human) user to access and to communicate with the system plays a crucial role. Channels and languages are central elements in the communication between agents in a multimedia system. An *agent* has an internal state, some goals and intentions, and the ability to communicate. A key attribute of communication is *modality*, the sensory system (auditory, visual, tactile, etc.) through which the communication occurs. A channel is an abstraction of a connection device that makes the communication between two agents happen. Communication needs to be meaningful. We call a communication an *interaction* if it results in a change of *state* of the other agent. A common language that can be written or read by the agents is a prerequisite for interaction. A *medium* is a set of co-ordinated channels, possibly spanning several modalities, which possesses a cross-channel language of interpretation. The user interacts with the system in form of *dialogues* to access content. During the usage of the system, the user creates a usually implicit *model* that describes her/his preferred way of interaction and that reflects goals and strategies.

4 Evaluation of Learning and Interaction Behaviour

4.1 Behaviour and Learning Interaction

Notions of interaction vary in different contexts: the learning process, interaction between human and computer, and communication in multimedia systems. Interaction in the learning context refers to activities of the learning process. Human-computer interaction separates the human knowledge-level and system-level aspects of interaction. Interaction in multimedia systems is based on technical notions addressing the communication of information and commands, which is reflected in a system's access log.

These different notions of interaction, however, can be integrated for educational software environments using models and languages. Activities that are central in learning style models are mapped to the services of the educational multimedia system. An evaluation of the interaction behaviour in the multimedia system can help us to determine the user's preferred learning style. In general, learning activities and educational multimedia services can be related. Patterns of learning behaviour, such as expressing study habits and strategies, can be associated with communication patterns.

4.2 Evaluation Objectives

The evaluation of the learning behaviour in interactive educational multimedia environments is the objective here. Data mining technology is the key tool in this endeavour. It allows us to extract the interaction behaviour of learners in form of patterns and rules from access logs. These rules and patterns of interaction behaviour shall be interpreted in the context of interactive learning. The formative evaluation objectives are *explanatory* (to understand learning processes and to validate designs) and *generative* (to create improved designs and utilise knowledge in e.g. adaptive and personalised systems). An assessment of the usability of the educational media is also possible.

Similar to Web mining for e-Commerce systems, where the analysis aims to support prediction and decision making for marketing and sales, the interpretation of mining results has to take place within an analytic model. Here, the model has to capture the context of learning and interactivity. Interactive educational multimedia as the educational environment need to be understood in order to interpret patterns and rules, and to integrate and utilise the results in evaluation and instructional design. Concrete evaluation goals include to identify novice or weak learners, to monitor individual learners or learner groups, to support usability analyses, or to carry out learner profiling (i.e. identifying learner characteristics) in order to determine a user's preferred learning style or to adapt a system to particular learner groups.

4.3 Languages and Models

The behaviour is a manifestation of the user's goals and strategies (Mullier, Hobbs, & Moore, 2001). The expression of interaction and learning behaviour through a language is the key tool in the evaluation of learner behaviour. *Multimedia interface languages* are based on actions such as mouse clicks or keyboard input. The user's interaction with a Web browser interface can be described in terms of specific interactions such as navigation and submission. Web logs are an expression of behaviour, reflecting the interaction of users with a Web-based multimedia system.

Interactivity models propose to distinguish different forms of learning interaction. Ohl (2001) introduces knowledge acquisition (an information request that is answered by a system output, such as navigation), knowledge production (information generation mostly in form of input to the system, such as submission), and operational acquisition (a combination of the first two). We capture the information flow and processing by an *interaction topology* consisting of nodes and arcs. Nodes represent states of the systems; arcs represent transitions between these states. State-transition notations are standard to describe the dynamics of computer-based systems. A state-transition-based *learning activity language* describes a learner's interaction with educational multimedia:

- *Nodes* represent the system's response including static output (text, image), dynamic output (animation), and input facilitating activity (form, editor). Nodes are described by attributes including modality, subject, and educational service.

- *Arcs* represent the activities offered to the user including links (navigation) and submission (action). Arcs are described by attributes including activity type (link, submit), action (the educational activity), action object (educational object), and source and target (nodes). For streamed media (audio, animations, video) interactions typically happen in form of actions such as start, stop, pause, forward, or rewind.

The goal – the evaluation of learning activity – is realised through the evaluation of access log expressions using mining techniques. A language describing navigation behaviour as represented in access logs and a language describing learning activities and interactions differ. In Web logs, activities are not explicitly recorded, only the resource requested by the action. However, learning activity expressions can be inferred from access log expressions. We can annotate nodes of access logs with activity-related arc types. This metadata language containing activity annotations is closer to the learning activity language. Web mining results based on annotated logs can be interpreted in a learning-specific analytic model. Sequential patterns identify learning tasks and activities, session classification determines predominant learning goals, and time series identify learning strategy and behaviour changes.

The learning activity language and its underlying concepts form the analytic model in which the access log is interpreted. The analytic model reflects the designers or evaluator’s understanding of interactivity, learning activities and educational multimedia. This language is often not explicitly formulated. However, explicit language and model would form an integrating component between evaluation and instructional design for educational multimedia.

5 Web Usage Mining for Educational Multimedia

Web usage mining extracts rules and patterns from access logs, see Figure 1. Different mining techniques support a variety of analysis and evaluation objectives.

5.1 Session Statistics

A session is a sequence of Web log entries that reflects the interaction behaviour of a learner in a period of active study. Some basic measures can help to answer questions about the investment of time for a given learning activity. Any of the results can be compared against the expectations of the instructional designer. Explicitly formulated expectations form part of the analytic model. There are other statistical measures that might result in useful insights. The total number of requests by interval or total numbers ranked by resource provide relevant information. These measures, however, give more an idea about ‘what’ resources are used than ‘how’ they are used.

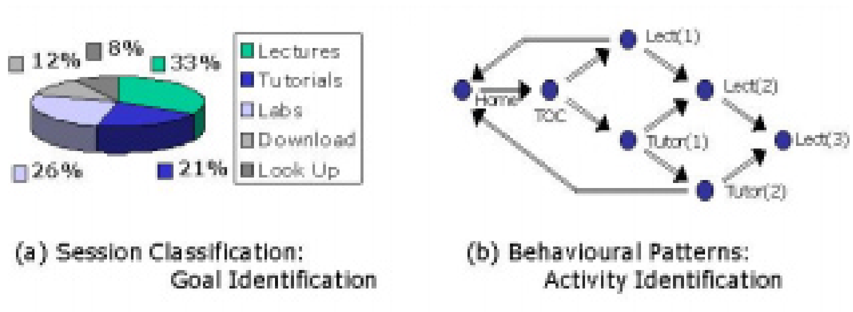

Figure 2. Data Mining Techniques.

5.2 Session Classification and Goal Identification

The objective of this technique is to extract the main goal(s) and higher-level tasks of the goal/task hierarchy from a session log (Donnellan, 2002). Typically, a learner focuses on one or two main activities in a session. Using a *classification* approach we can identify the main learning objective(s) by looking at the predominant types of interaction with the system, Figure 2(a).

The media resources of a course Web site can be classified, i.e. classes $C_1, \ldots, C_N$ are created where each class C_i is a set $\{U_{i1}, \ldots, U_{iM}\}$ of URLs. Each class C_i is associated to a type of system-level interaction that facilitates a particular knowledge-level learning activity, such as attending a virtual lecture or working on virtual lab exercises. If a learner spends substantial session time on a particular activity, then this activity is a manifestation of a particular goal. The requests of pages of the individual classes are counted.

For each session, a *ranking* $C_{i1} \leq \ldots \leq C_{iN}$ of main learning goal(s) represented by learning activity classes is produced based on the number of requests for each class, which gives us some insight into the goal/task hierarchy and the importance of some of the goals and tasks. This can be generalised for all sessions of a learner or for groups of learners. This technique can be used in an iterative evaluation process. Initial classifications might turn out too unspecific and can be refined into more fine-granular classifications, identifying more specific activities, tasks, and goals; thus providing more detailed and meaningful analysis results.

5.3 Behavioural Patterns and Activity Identification

The goal identification is a tool on an abstract level. Often, however, a closer look at interactions at a lower, fine-granular level is necessary in order to investigate learning activities in detail. The objective of this mining technique is to extract behavioural interaction patterns from the log file. Irrelevant activities – students might lookup other pages, even leave the system temporarily – can be discarded. The filtered sequences are candidates for *sequential patterns*. In order to find out what patterns learners follow, the sequences are subjected to some threshold control – another filter to discard too uncommon ones.

Behavioural patterns encompass more than sequences – learners repeat elements, choose between options, or work on several course elements in parallel. A model of the course topology – navigation infrastructure and interactive elements abstracted by nodes and arcs – underlies behavioural patterns, Figure 2(b). A *behavioural pattern* is an expression of a learning activity language that describes potential or actual learning as interactions with an educational multimedia system.

$$(\mathit{attendLect(1)} \mid \mathit{attendTutor(1)}^{+}) \; ; \; (\mathit{attendLect(2)} \mid \mathit{attendTutor(2)}^{+}) \; ; \; \ldots$$

In this example, *attendLect(1)* and *attendTutor(1)* are activities. The expression specifies that the learner can use either lectures or tutorials (the | -operator). The tutorial might be attended repeatedly (the $^{+}$ -operator), before the next lesson is looked at (the ; -operator). These expressions can also be represented in a graphical form, see Figure 2(b).

These behavioural patterns can be a reflection of the instructor's intended use or an abstraction of the learner's behaviour. The learning activity language to describe these patterns is an integral part of the analysis model that is used to interpret mining results. Therefore, we need to relate these behavioural patterns with the sequential patterns extracted from the Web log. As explained in Section 4.3, activities can be associated with the transitions between the nodes (URLs) reflected in the log. More advanced results can be achieved if for example the time spent on each activity and other properties are included in the evaluation (Xu, 2003).

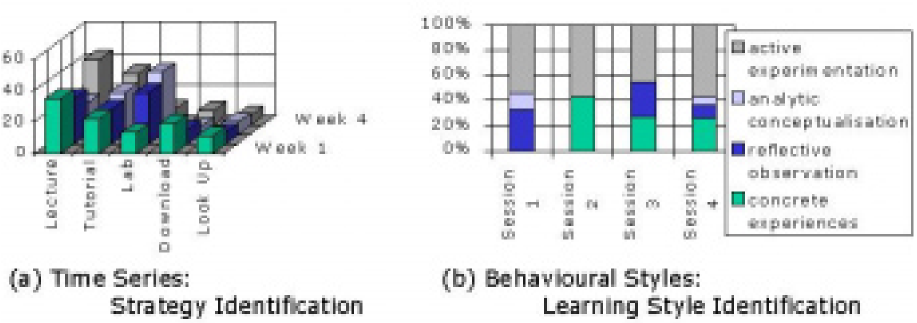

Figure 3. Data Mining Techniques.

5.4 Time Series and Strategy Identification

Time series are sequences of measurements over a period of time, Figure 3(a). These measurements can include results from any of the mining techniques. The purpose is the detection of change in learning behaviour, which is often a reflection of the overall learning strategy over the duration of a course. This is important for two reasons. Firstly, change might be intended by the instructional designer and the actual occurrence of change needs to be verified. Secondly, unexpected changes need to be detected. Time series evaluations allow the detection and constant monitoring of student activity changes.

An example for the first case is an evaluation of scaffolding features through behavioural pattern analysis. Fading use of scaffolds – features that support students in becoming self-reliant and competent in a topic – is an essential characteristic that is expected to happen in an effective scaffolding implementation. Apart from behaviour change, changes of learning strategies have also been observed in the case study. Early patterns often show single-goal use, but later patterns show a concurrent, integrated usage of different educational services. Time series of usage patterns can illustrate the evolution of student learning from Web logs.

5.5 Behavioural Styles and Learning Style Identification

Learning style models address a learner's preferred way of learning. One category of models has activities of information processing based on a learner's interactions with the educational environment at its centre – an example is Kolb's Learning Style Inventory (Kolb, 1984). Another category is sensory perception oriented – related to modalities in multimedia systems.

Kolb, for instance, proposes a theory of experimental learning – suitable for capturing learning activities – that involves four activity dimensions: concrete experiences, reflective observation, analytic conceptualisation, and active experimentation. Kolb derives four types of learners based on their preferred activities. Pages of the database course were classified according to activities, for instance, concrete experiences: tutorial animations, or reflective observation: lecture material giving context and concepts. These can be associated to Kolb's learner types, see Figure 3(b). Using classification yields a ranking of classes and, consequently, an indication of the preferred learning style.

5.6 A Case Study – Experience in a Web-based Course Environment

The mining techniques have been used extensively in a virtual course environment (Pahl & Donnellan, 2002). The virtual course – a database course taught to third-level computing students – focuses on practical aspects of database engineering. Various types of media implemented using Web technologies facilitate different forms of interaction with the course content. Several learning activities including attending lectures and working in tutorials and labs are supported through an integration of different interactive educational services. The *lectures* consist of combined audio-visual material – the recorded lecturer's speech synchronised with the presentation of Web pages. The audio stream is fully controlled by the student. The tutorial and lab services support active learning. *Tutorials* are supported by various forms of animations based on HTML and flash technology – again with a high degree of control of presentation by the student. *Lab* features facilitate interaction with software tools. Various styles ranging from form fill to direct manipulation are provided that allow students to acquire and train skills. Additional resource centres and self-assessment features complement the core services.

The *integration* of different educational services, which themselves comprise a variety of interaction styles and media, enables a number of possible learning strategies and activities for the students. The evaluation of the student's learning behaviour based on the actual course usage is therefore of paramount importance. The course has been evaluated using freeware data analysers for basic statistical analyses and custom-built mining tools for the advanced aspects (Donnellan, 2002; Xu, 2003). The use of usage mining technology has benefited both learners and instructors alike.

Web mining has been used in different styles.

- In a *predictive* style to confirm expectations and validate designs. Learning styles and strategies and their change over time were analysed. An example is the design of lab and scaffolding usage support, which has been validated through predicted learning behaviour based on mining data.
- In a *generative* style to improve the design. The instructor's expectations of student learning behaviour were expressed in a learning activity language and compared with actual behaviour. The navigation features and the course topology have been gradually improved using this technique.
- In an *explanatory* style to understand student learning in a novel environment. Mining techniques have been used to clarify the understanding of the learner's goals and task hierarchies. Mining techniques have been used to investigate the sequential and concurrent use of features. This analysis formed the starting point for the design of a multi-feature learning environment.

Education-specific mining techniques – see Figures 2 and 3 – have helped to improve the instructional design and also to confirm delivery-related decisions that were made. It has supported instructors in providing quality instruction and in achieving a better learning experience for learners.

6 Discussion and Conclusions

Customer relationship management is a central aim in the e-commerce environment. We suggest to make learner relationship management an objective for e-learning. As evaluating and understanding customer needs and customer behaviour, understanding the learner and learning processes is essential for instruction. More insight into learner needs and behaviour through evaluation supports the instructor in optimising and improving instructional design and the learning experience – in particular for novel and evolving learning environments such as the Web.

The benefits of data mining technology in discovering latent knowledge make it a convincing solution for the observation-based evaluation of learner behaviour in feature-rich educational environments in particular. The comparison with classical forms of evaluation, which are usually error prone, intrusive and difficult to automate, emphasises the benefits. Further improvements could be achieved by widening the evaluation basis.

Mullier, Hobbs, and Moore (2002) suggest content mining – the automated analysis of content. Moreover, combinations with classical evaluation techniques are possible (Oliver, 2000).

Interaction is central in active learning. We have embedded this learning-oriented notion of interaction into a more technical view of interaction in multimedia systems, based on the hypothesis that interaction behaviour is a reflection of learning goals and strategies. Data mining is a technology suitable to extract different aspects of a learner's interaction with educational media from access logs. Mining results can be interpreted in analytic educational models of interaction and learning to further our understanding or to use learner-specific context data in personalised learning and adaptive educational multimedia.

Models and languages allow us to express goals and tasks as learning concepts, to capture interaction behaviour in educational multimedia systems, and to relate and integrate these descriptions. More work has to be invested into languages and models in order to integrate design and evaluation. The presented results are a first step towards an incremental methodology for instructional designers based on design and evaluation iterations that reflect evaluation results in the instructional design. Formative evaluation is the strategy toward better understanding and improved design, but only if we integrate pedagogical and technical concepts of activity and form a coherent analytic model, we are able to successfully improve instructional design for educational multimedia through Web and data mining.

Acknowledgements

The research described here has been supported by the Dublin City University Teaching and Learning Fund. The author wishes to thank Lei Xu and Dave Donnellan, who implemented most parts of the mining techniques described here.

References

Agrawal, R. & Srikant, R. (1995). Mining Sequential Patterns. *Proc. 11th International Conference on Data Engineering ICDE, Taipei, Taiwan*.

Chang, G., Healey, M.J., McHugh, J.A.M., & Wang, J.T.L. (2001). *Mining the World Wide Web - An Information Search Approach*. Boston: Kluwer Academic Publishers.

Cooley, R., Tan, P.-N., & Srivastava, J. (1999). Discovery of Interesting Usage Patterns from Web Data. *Proc. WEBKDD’99*, 163-182. London: Springer-Verlag.

Dewey, J. (1916). *Democracy and Education*. New York: MacMillan.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human-Computer Interaction*. London: Prentice Hall.

Donnellan, D. (2002) .*User Session Classification Tool for the Analysis of Web Server Logs*. M.Sc. Dissertation. Dublin City University. School of Computing.

Elsom-Cook, M. (2001). *Principles of Interactive Multimedia*. London: McGraw-Hill.

Hirami, A. (2002). The Design and Sequencing of e-Learning Interactions: A Grounded Approach. *International Journal on E-Learning*. 1(1).

Jonassen, D.H. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology*, 34(4), 34-37.

Jonassen, D.H., & Mandl, H. (eds) (1990). *Designing Hypermedia for Learning*. Berlin: Springer-Verlag.

Kolb, D.A. (1984). *Experiential Learning*. New Jersey: Prentice Hall.

Laurel, B. (1993). *Computers as theatre (2nd Ed.)*. New York: Addison-Wesley.

Moore, M.G. (1992). Three types of interaction. *The American Journal of Distance Education*, 3(2), 1-6.

Mullier, D., Hobbs, D., & Moore, D. (2002). Identifying and Using Hypermedia Browsing Patterns. *Journal of Educational Multimedia and Hypermedia*, 11(1), 31-50.

Norman, D.A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

Norman, K.L. (1998). Collaborative interactions in support of learning: Models, metaphors, and management. In Hazemi, R., Wilbur, S., & Hailes, S. (eds). *The Digital University: Reinventing the Academy*. London: Springer-Verlag.

Northrup, P. (2001). A Framework for Designing Interactivity into Web-based Instruction. *Educational Technology*, 41(2), 31-39.

Ohl, T.M. (2001). An Interaction-Centric Learning Model. *Journal of Educational Multimedia and Hypermedia*, 10(4), 311-332.

Oliver, M. (ed.) (2000). Special Issue on Evaluation of Learning Technology. *Educational Technology & Society*, 3(4).

Pahl, C. & Donnellan, D. (2002). Data Mining for the Evaluation of Web-based Teaching and Learning Environments. *Proc. E-Learn 2002 World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*. AACE.

Ravenscroft, A., Tait, K., & Hughes, I. (1998). Beyond the Media: Knowledge Level Interaction and Guided Integration for CBL Systems. *Computers and Education*, 30 (1/2), 49-56.

Sims, R. (1997). Interactive learning as "emerging" technology: A reassessment of interactive and instructional design strategies. *Australian Journal of Educational Technology*, 13(1), 68-84.

Weston, T.J. & Barker, L. (2001). Designing, Implementing, and Evaluating Web-Based Learning Modules for University Students. *Educational Technology*, 41(4),15-22.

Xu, L. (2003). *Analysis of Behavioural Patterns using Web Mining*. M.Sc. Dissertation. Dublin City University. School of Computing.

Zaïane, O.R. (2001). Web Usage Mining for a Better Web-based Learning Environment. *Proc. Conference on Advanced Technology for Education CATE'2001*, 60-64.

Zaïane, O.R. (2002). Building a Recommender Agent for e-Learning Systems. *Proc. 7th International Conference on Computers in Education ICCE'02*, 55-59. AACE.