PROBLEM-BASED LEARNING IN DATA MINING

Pirjo Moen
Department of Computer Science
P.O. Box 68
FI-00014 University of Helsinki
pirjo.moen@cs.helsinki.fi
http://www.cs.helsinki.fi/pirjo.moen

ABSTRACT

Good teaching is teaching where the teaching method and assessment methods are aligned with the learning activities stated in objectives of the course. Typically such teaching is also student-focused. One example of such a teaching method is problem-based learning.

Problem-based learning has been widely studied and applied in many fields. In this paper we describe one example of how problem-based learning has been applied in teaching of data mining. We first describe the structure of a course of methods for data mining. Then we analyse the course both from a teacher's and from a students' perspective. We also present some ideas on how this course could still be improved in the future.

Keywords

Problem-based learning, Data mining

1. INTRODUCTION

Good teaching is aligned teaching, i.e., teaching where the teaching method and assessment methods support the learning activities stated in objectives of the course [2, 11]. It is also characteristic of good teaching that it supports deep approach to learning and it is based on students' activity. Other factors related to good teaching and learning are the social aspects of learning, e.g., working in a group, and giving the students possibilities for connecting the new information to the things they knew before. Good teaching is, therefore, clearly student-focused [11].

An example of a teaching method that can fulfil the requirements of good teaching above is *problem-based learning*. This teaching and learning method has been widely studied and applied in many fields, e.g., in computer science, jurisprudence, medicine, veterinary medicine, and pharmacy [7]. In computer science this method has been applied, for instance, in teaching of computer graphics [1], artificial intelligence [3], programming [1, 4] and operating systems [4].

In this paper we describe how problem-based learning has been applied in a course of methods for *data mining*. This course, called Data Mining Methods, is an advanced level course that is a part of the Master of Science degree at the Department of Computer Science at the University of Helsinki, Finland. At the moment it is one of the two compulsory courses of the specialisation area of data management. The aim of the course is to introduce to the students the basic data mining concepts and methods, and the knowledge discovery process. This course was lectured in the presented form in Spring 2004 and 2005.

The rest of the paper is organized as follows. First, in Chapter 2 we briefly describe problem-based learning. Then, in Chapter 3 we present how the course Data Mining Methods was organized and how the problem-based learning method was applied in this course. In Chapter 4, we analyse the course both from a teacher's and from a students' perspective. After that, in Chapter 5 we present some ideas on how this course could be improved and, thus, student learning hopefully be enhanced. Chapter 6 is a short conclusion.

2. PROBLEM-BASED LEARNING

In problem-based learning the students are encouraged to work as experts and tutors for their own learning process [5]. Problem-based learning is a process where a student has much more responsibility on his/her own learning than in more traditional teaching methods. Therefore, problem-based learning is a good example of a student-focused teaching and learning method. Other typical characteristics of problem-based learning are combination of theory and practice, working in a group and development of students' collaboration and communication skills, emphasizing the learning process and development of the learning skills, continuous self-assessment of gained knowledge

and skills, and considering problems across subject boundaries [2, 4, 7].

Problem-based learning is based on assignments that should be as near to the real-life problems as possible [4, 5, 7]. These assignments should be so large and difficult that they cannot be solved based on the previous knowledge of the students. In fact, the aim of this method is not to find right solutions to the given problems, but instead, to try deeply understand the problems and to look them from different viewpoints [7]. In many cases, there are not even any single right or wrong solutions to the problems. And when the problems are large, open-ended and demanding enough, it is more useful to try to solve them in a group, instead of working with them alone.

Typically the problem-based learning consists of seven steps [5, 7]. These steps are

1. setting up the context and defining unclear concepts;
2. presenting research problems;
3. creating working theories and hypotheses (brainstorm);
4. critical evaluation of the theories;
5. defining learning objectives for this problem;
6. searching for deepening knowledge;
7. discussion in a group and distribution of expertise.

The first five steps are completed in a group in the session where the problem is presented to the students. During the step six, the students work independently, without the teacher or even the fellow students. After that, the students discuss again in a group and look at what they have learned and evaluated that with respect to the learning objectives they had defined earlier. These steps are just one way of defining the problem-based learning process, and they do not have to be followed exactly when problem-based learning is used [5].

As mentioned earlier, problem-based learning has been widely studied and applied in many fields, including computer science. Actually, it is a very suitable method for teaching computer science, because this field is by nature problem-based [1, 4]. Development and changes in computer science are very fast and continuous, and in order to keep up with the field, every computer scientist must learn new things all the time. Working in project groups is also very typical for computer science industry, and, thus, the students need to have good collaboration and communication skills when they graduate. Problems in computer science also often cross subject boundaries. Therefore, using problem-based learning can give computer scientist many valuable skills for solving the real-life problems.

3. DESCRIPTION OF THE COURSE

Knowledge discovery from databases or data mining is an area in computer science that aims to find methods for discovering interesting patterns or exceptions from large amounts of data. In the course Data Mining Methods the main aim is that the students should become acquainted with the basic concepts of data mining, learn what the main types of methods used in data mining are and when they can be applied, and what is meant by the knowledge discovery process. This course contains both theoretical and practical material. In the course we have lately used problem-based learning, which means that, in contrast to the traditional courses, the students in this course worked in small groups and tried to solve together some data mining problems. In this form, the course was lectured in Spring term 2004 and 2005.

The course consisted of six modules, each of which was two weeks long and had a specific theme [8, 9]. Lectures of the course were usually given on the first week of each module, except during the first module, when the lectures were given on both weeks. The lectures were mainly based on the book *Data Mining – Concepts and Techniques* [6]. During the first five modules, we had two exercise sessions per week (2*2 hours) where the student group gathered for their meetings and they could also ask from the teacher advice on how they should proceed with their current problem. During the last module these exercise sessions were replaced by a two-hour poster session, where each group of students presented a poster of the problems and solutions they had worked with during the course. In these presentations we used a technique called gallery walking [7]. At the end of the course there was a course exam.

During the first exercise session the students formed groups of 2-4 persons. These groups worked with one problem per each module, i.e., they worked with five problems in total. In the first exercise session of each module the groups were given an assignment on the theme of the module, and from the second module on, the assignments of the previous module were also discussed together in these sessions. Otherwise, the teacher followed the discussions of the groups and gave advice on how the groups should proceed with their work. Sometimes the groups had meetings also between the exercise sessions, as they were recommended.

The assessment of the course consisted of two parts: the reports of group work and the final exam. From each report, the students were able to get 4-6 points, and in addition to that at maximum three points for the poster session (the poster and the presentation of it). That is, the maximum number of points from the exercises was 30. Similarly, the maximum number of points from the final exam

was 30 points. The course was completed if the student was able to collect 30 or more points, so that at least 15 of them came from the final exam and at least 10 of them from the exercises.

In the assignments (problems) we asked each group of students to form a consultant office that specialised in data mining. The consultant offices were supposed to help their customers to decide whether data mining was an appropriate method for analysing their large data sets. The groups could choose one data set out of three to represent their customer's data. At the end of each module, the groups submitted a written report in which they described, in their own words, the data mining methods of the module, what kind of experiments they had done with the customer's data using such methods, and what the main results were. In each report they were also supposed to include a short learning diary describing what they had learned during that module, which things had been easy and which difficult, and what the learned things meant to them.

In the final exam there was one question per module (or theme). Some of these questions were theoretical, asking the students to explain briefly some basic concepts of the course. Typically, the students were also asked to give some real-life examples of these concepts. Another type of questions was formed by questions where the students were given a small data set and they were asked to apply some given mining method to it.

4. ANALYSIS OF THE COURSE

In the previous chapter we describe the structure of the course Data Mining Methods, and how problem-based learning was applied in the course. Now we look at what was good in the course and what things could still be improved. We consider these issues both from the teacher's and the students' perspective.

4.1 Teacher's perspective

One of the pitfalls of problem-based learning is that teachers are not familiar enough with theoretical background and the principles of the method [7]. When we started to plan the Data Mining Methods course, we studied the pedagogical literature on the method, but, unfortunately, we had no previous experience on applying the method. Right in the beginning of the course, we explained to the students the basic features of the problem-based learning method, and how the method was to be applied in this course. One of the mistakes that we made at this point – and later as well - was that we trusted that, after this introduction, the students know how the method works, and we did not follow their work in the groups as carefully as we should have done. Another problem that we noticed was that obviously the role of the teacher as a tutor of the groups was, to some extent, unclear to both students and teachers.

As mentioned before, our course consisted of six modules with the lectures given right at the beginning of each module. From a teacher's standpoint, and as the literature also shows, this was not a good decision: in problem-based learning the lectures should really be given first when the students are deepening their knowledge on the problem, i.e., first after the students have had time to think about the problem given to them and to formulate some kind of questions or even solutions to the problem [5]. The practices used in the course just guided the students to read only the course book and the lecture slides, and not to do any information searching of their own. Otherwise, the structure of the course seemed to be good, especially the poster session with gallery walking turned out to be very stimulating.

Group dynamics in the course were mostly good, and the students really worked as a team. Unfortunately, there were some groups that were not that successful in their collaboration. One reason for this might have been the differences in the learning objectives that the students in these groups had. Some of these groups were able to solve their problems, and to complete the course with good results, while some groups could hardly finish their reports. In general, nearly all the groups had difficulties in finding time for their meeting, despite the organised weekly exercise sessions.

In a course where a problem-based learning method is used, a teacher is able to use most of his/her time interacting and tutoring the students, instead of preparing lectures [1]. Such interactions are usually very stimulating and rewarding, and the teacher can learn to know his/her students better and track their learning. The teacher can also easily form a picture in his/her mind of what the students can, or what kind of learners they are. What surprised us, was that a student who was active in discussions and in group work and seemed to have learned a lot, could totally fail the final exam of the course. Probably, in these cases, the student had not really understood all the parts of the problem.

About 50 students attended the course Data Mining Methods during the two times when it was lectured in this form. Of these students, almost 93 pro cent passed the course, which is more than on our traditional courses. The failure or poor performance in the final exam, still, indicated that some students have problems in learning. Some of these problems might be due to the assessment methods used in the course. One possibility might, for example, be that the questions in the exam were dependent on each other, as was suggested in [3]. In our course, the exam questions were, however, totally independent on each other, so that cannot explain

the phenomenon. A more plausible explanation is, therefore, that active participation in discussions and in group work does not necessarily guarantee that the student has really understood the subject or has really deeply thought of it [3]. Even though we tried to design the problems in this course so that they could not be divided into sub-problems, the students might still have managed somehow to do that. In such a case, each student is, of course, responsible for carefully studying the parts of the work done by the other members of the group.

One problem in this course that we noticed and had not expected was the fact that some students did not have enough background knowledge. Some of them had not even completed all the compulsory intermediate-level courses in computer science. In addition to this, it turned out later, that some of the intermediate-level courses did not even include all the issues that we had believed them to cover. One possible solution to this problem is to give the students in the future a clear list of required background material they are expected to know and to advise them to study that material on their own. A list of the possibly useful materials and information sources would also be very useful for the students [3, 7].

4.2 Students' perspective

Problem-based learning as a teaching method was totally new to most of the students attending this course. Learning diaries and discussions with the students made it clear, that many students were quite unsure, confused, and even frightened at the beginning of the course. The reason for this was that they did not really know how this learning method works and what they were expected to do. Gradually, most of these students understood why this method was used in this course, and they were really happy about all the things they had been able to learn.

Some of the students were, however, angry and bitter throughout the course. According to them, they were not able to learn anything, because nothing was taught, or the things that were taught, were taught in a wrong way. It showed that these students were very used to the practises used in some other courses, and they could not see how problem-based learning could enhance their learning. They would have preferred to attend a very traditional course, where a teacher gives them via the lectures all the information needed for passing the course.

As mentioned before, the group dynamics worked quite well, and the students were mainly satisfied with the group work. According to them, working as a group made them more committed to the course, which seems to be characteristic for this kind of courses [1, 3]. This was also shown by the fact that, after the first module, no one dropped out of the course.

In problem-based learning, one of the main ideas is that the problems should be clear, but still open-ended and large enough, so that no one can solve them based to his/her previous knowledge. The problems considered on the course Data Mining Methods were made according to this principle. However, some students in the course thought that the problems were too large and too open-ended, and they were quite annoyed that there were no right answers to the problems. Obviously, they did not understand that finding right (perhaps even non-existing) solutions to the problems was not important, and that it was more important to study the problem deeply from different perspectives.

5. DEVELOPMENT OF THE COURSE

In the previous chapter we noticed that the problem-based learning method worked well in many ways in the Data Mining Methods course. Still, there are ways to improve the course further. One of them is to give for the students more complete written instructions on problem-based learning and how it is applied in the course. Another way to enhance learning and improve the course is to use more time at the beginning of the course for the introduction of the teaching method and for carefully explaining, why this method has been chosen for this course and why it is so useful [7].

An important part of the information given to the students at the beginning of the course is the learning objectives of the course. In the Data Mining Methods course, one of the main aims is that students learn to know and understand what is meant by a knowledge discovery process and what kinds of phases there are in this process. Another aim of the course is that the students form their own opinions and knowledge on what kind of pre-processing can and should be done to the data when different data mining methods are applied. The third, and perhaps the most important, learning objective of this course is that the students learn to know what kinds of data mining methods there are, and that they can evaluate when these methods can or should be used in analysis of the data sets. These learning objectives have existed in the course all the time, but they might not have become clear enough to the students attending the course. Therefore, we are going to explicitly write down these objectives for the next time the course is held.

Yet another change we are going to make is to change the time and the role of the lectures in this course. We could, of course, leave the lectures totally out of the program, but, at the moment, that seems to be a bit too radical. Instead, we plan now to have the lectures in the second week of each module, i.e., in the phase when the students are

already deepening their knowledge of the problem at hand. We hope that this encourages the students to actively search for more information, and thus to enhance their learning.

To support of each individual we are also planning to improve the questions in the final exam. This is to be done so that the questions will be more similar to the problems used in the course, which is supposed to be essential for questions in an exam of a problem-based course [10]. Learning diaries will also be individual in the future.

In addition to the learning objectives concerning the subject of the course, in our case methods for data mining, problem-based learning should assist the students in developing their learning, collaboration and communication skills. The students should also have a possibility to demonstrate this kind of learning during the course [10]. Therefore, we are going to add peer and self-assessments to this course, which are natural ways of assessing learning process, and collaboration and communication skills [2, 3, 10].

6. CONCLUSIONS

In this paper we have presented a problem-based course for teaching methods for data mining held at the Department of Computer Science at the University of Helsinki, Finland. We showed that problem-based learning is a suitable and very natural teaching method for this kind of a course. Still, we were able to identify several practises of the course that could be improved in the future. Most important of these were the need of proper instructions on problem-based learning and how this method is applied in the course, and the need of writing down the exact learning objectives of the course. These changes we shall make in the near future.

7. REFERENCES

[1] Baldwin D., Discovery learning in computer science. *ACM SIGCSE Bulletin* **28 (1)**, 222-226 (1996).

[2] Biggs J., *Teaching for Quality Learning at University.* 2nd edition. The Society for Research into Higher Education (2003).

[3] Cavedon L., Harland J. and Padgham L., Problem Based Learning with Technological Support in an AI Subject: Description and Evaluation. *Proceedings of the 2nd Australasian Conference on Computer Science Education* (ACSE'97), 191-200 (1997).

[4] Ellis A., Carswell L., Bernat A., Deveaux D., Frison P., Meisalo V., Meyer V., Nulden U., Rugelj J. and Tarhio, J., Resources, Tools, and Techniques for Problem Based Learning in Computing. *ACM SIGCUE Outlook* **26 (4)**, 41-56 (1998).

[5] Hakkarainen K., Lonka K. and Lipponen, L., *Tutkiva oppiminen – Järki, tunteet ja kulttuuri oppimisen sytyttäjinä.* WSOY (2004).

[6] Han J. and Kamber M., *Data Mining – Concepts and Techniques*. Morgan Kaufmann (2001).

[7] Lindblom-Ylänne S. and Nevgi, A. (eds), *Yliopisto- ja korkeakouluopettajan käsikirja.* WSOY (2002).

[8] Moen P., Data mining methods – Spring 2004. http://www.cs.helsinki.fi/pirjo.moen/tilome/tilomeK04.html.

[9] Moen P., Data mining methods- Spring 2005. http://www.cs.helsinki.fi/pirjo.moen/tilome/tilomeK05.html.

[10] Rawnsley K., Spaziani R. and Rangachari, P. K., Evaluation in problem-based course: contrasting view of students and teacher. *Probe* **12**, 9-14 (1994).

[11] Trigwell K., Judging university teaching. *The International Journal For Academic Development* **6 (1)**, 65-73 (2001).