LEARNING BY DOING AND LEARNING WHEN DOING

Dovetailing E-Learning and Decision Support with a Data Mining Tutor

Klaus P. Jantke, Steffen Lange
German Research Center for Artificial Intelligence, Saarbrücken, Germany

Gunter Grieser, Peter Grigoriev
Technical University of Darmstadt, Dept. Informatics, Darmstadt, Germany

Bernhard Thalheim
Christian-Albrechts-University of Kiel, Dept. Informatics, Kiel, Germany

Bernd Tschiedel
Technical University of Cottbus, Dept. Informatics, Cottbus, Germany

Keywords: E-Learning, Knowledge Discovery, Data Mining, Decision Support, Learning on Demand

Abstract: In this paper, e-learning meets decision support in enterprises' business practice. This presentation is based on an on-line e-learning system named DaMiT for the domain of knowledge discovery and data mining (see http://damit.dfki.de). The DaMiT system has primarily been developed for technology enhanced learning in German academia. It is now on the cusp of entering training on demand in enterprises. Simple stand-alone e-learning seems quite unrealistic and does not meet the needs of industry. It is very unlikely that employees take a detour to study theory of whatever sort. More likely, they are willing to engage in studies whenever the need derives directly from their practical work. In those cases, they might even be willing to dive into theory. How to dovetail e-learning and enterprise business applications, such that both sides benefit from it?

1 INTRODUCTION

There is no doubt at all that technology enhanced learning is going to change education on all levels ranging from schools over universities to professionnal training and lifelong learning. The process is boosted by the Internet in pervading the world.

The recently observable progress in the area named e-learning is enormous and ranges from a flood of content (For illustration, the German Federal Ministry for Education and Research, BMBF, has put about 200 Mio. Euro into 100 joint projects to develop content for academic e-learning. Another 200 Mio. Euro went into schools and professional education, all this within only 3 years.) to technological innovations. We are all on the cusp of inventing truly adaptive e-learning systems, based on deep learner modelling, expressive XML-based content representation and flexible, attractive and appealing generation and presentation technologies.

There are still a number of open problems also in technology, but the R&D community is very active.

In industry, however, one observes an obvious reluctance. Employees tend to restrain from getting involved in extra activities. Moreover, the management frequently has understandable reservations about introducing another software system and further diversifying the IT infrastructure.

This situation bears abundant evidence for the need of truly integrating e-learning into business processes and IT infrastructures of enterprises.

Last but not least, the questions under discussion are relevant to universities and other academic institutions when pondering about marketing potentials.

P. Jantke K., Lange S., Grieser G., Grigoriev P., Thalheim B. and Tschiedel B. (2004).
LEARNING BY DOING AND LEARNING WHEN DOING - Dovetailing E-Learning and Decision Support with a Data Mining Tutor.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 238-241
DOI: 10.5220/0002623602380241

2 ALMOST AN EXCUSE

Due to the necessity to reduce the submission from 8 down to 4 pages, the authors refrain from any deeper discussion of what data mining and decision support are about. However, DaMiT is an e-learning system facing the subject of data mining. Studies in this area do require intensive **learning by doing**, and when decision support in enterprises is ongoing, the DaMiT system is offering a framework of integrated learning on demand called **learning when doing**.

3 ACTIVE LEARNING IN DaMiT

This chapter contains a detailed discussion of **learning by doing** in the DaMiT system. Chapter 4 is showing how to exploit the doing-oriented features of the system for **learning when doing**, e.g., in industrial settings, in governmental working environments or in research institutes. To bridge this gap is the aim of the present paper.

3.1 Observational Learning

Data mining may be considered both a science and an art. When practicing the art of data mining, a quite substantial amount of the underlying knowledge is implicit. But how to transmit implicit knowledge by means of e-learning? This is a particularly tough question if the teachers are not always aware of the knowledge they are propagating when being engaged in teaching. Sometimes, one says you *just need to get some feeling about it*.

The problem of implicit knowledge is even more important when the domain is a rather young one and results are not matured, established publications do not yet exist and teachers are not experienced, as it usually applies to data mining.

The DaMiT system is equipped with several “playgrounds” where learners can experience those phenomena which are rather difficult to deal with explicitly. Learners experience different phenomena like, e.g., how very small changes in the input data result in enormous changes of the classifier induced or, alternatively, when substantial changes to the data do not change the hypothesized classifier at all.

It is surely one of the highlights in education when learners are able to pose interesting problems to their teachers. Figure 1 displays an applet where a learner can generate decision problems to be solved by the DaMiT system in generating decision trees over regular patterns. The learners provide the input data, i.e. positive and negative examples, and the system generates a certain decision tree with regular patterns serving as tests in the nodes of the tree. The learners can inspect the generated tree and, then, can modify the posed learning problem according to their ideas of how to make the learning task more easy or more difficult to the system.

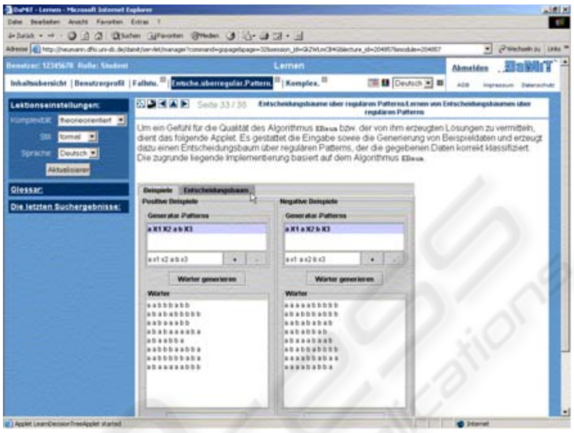

Figure 1: An Applet for Posing Tasks to the System

Observational learning of this type can not easily be substituted by other learning forms. Data mining always contains a phase of exploration.

3.2 Experiencing True Data Mining

There is not much hope for learning to swim or to ride a bicycle when sitting on a sofa, only. Quite analogously, there is not much hope for learning data mining by reading text books or texts of some web-based e-learning system like DaMiT, only. You need to do it, and you need to do it properly.

The DaMiT system contains case studies as well as what we call *competitive exercises*. Those exercises are of the quality of practical data mining problems. There is a continuous competition among all learners in finding better and better solutions to these problems.

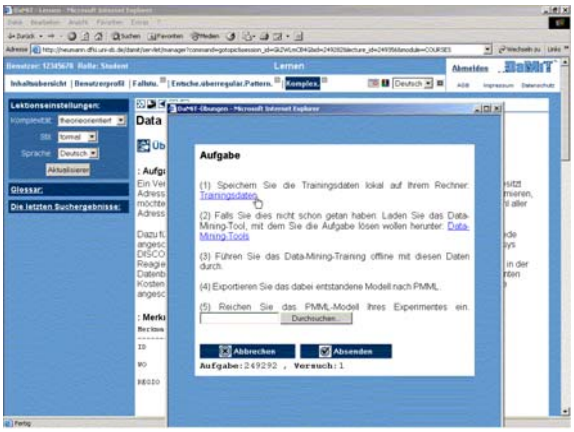

Figure 2: A Competitive Exercise in DaMiT

4 LEARNING ON DEMAND

When in practice problems do arise which may be explained or interpreted over large and usually distributed data, the essence is rarely sufficiently understood in the very beginning. Symptoms are recognized, but a useful diagnosis may take some time and may be laborious.

For illustration, an enterprise's management may recognize a growing number of customers cancelling their business relations with that enterprise. A first self-evident management decision might be to ask somebody to look into the individual data and find out the reasons. If this fails, what to do next?

Even if no pressing problems urge the management to inspect larger data bases, certain desires for cost reduction may lead to the wish of understanding relations not properly understood so far. For instance, if a mailing action in direct marketing shall be more focussed than it used to be before, one should find out which customers are very likely to respond and which are not. Again, a self-evident management decision might be to ask somebody to look into the data and tell which customers are to be addressed. If this fails, what to do next?

To have an appropriate e-learning system at the management's fingertips may help a lot. You can get consulting about the general problem you are facing, you can get knowledge about approaches and technologies, you can get tools for attacking your problem, and, finally, you can get support in evaluating your own solution to the problem you have.

In a system like DaMiT, as seen above, you can find problems similar to the one you are facing. And you can get all this for free, because a large amount of the e-learning content is open to the public.

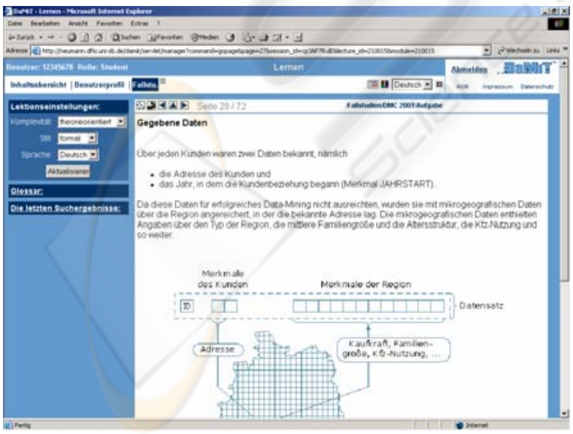

Figure 3: A Case in Direct Marketing Optimization

More generally speaking, with a system like DaMiT one can get consulting about the characteristics of a problem, about basic variants and crucial details to be considered, and about ways of how to go forward towards a solution. This means already learning when doing.

Assume that a problem like that of finding those customers which are likely to respond to a mailing activity is understood as a classification problem. Let us further assume that the general principles of decision tree induction are understood and believed to be helpful. (If not, consult the system and *learn* more about this area.) Then it is a management decision to go for generating a decision tree classifier over the own data base.

In that case, data understanding and data preparation are inevitable steps. There is no hope at all to take your data as they are and start learning any useful classifier. In practice, this problem is generally awkwardly underestimated. In enterprises, one may study the lessons and try the tools for data understanding. Doing so means learning when doing.

If the tool has been chosen – we take QuDA, a tool developed at TU Darmstadt, in the sequel, for illustration – and the data are prepared, one can get involved in the laborious process of interactively generating a classifier.

Normally, one comes up with a first classifier, inspects it and returns to the generation process.

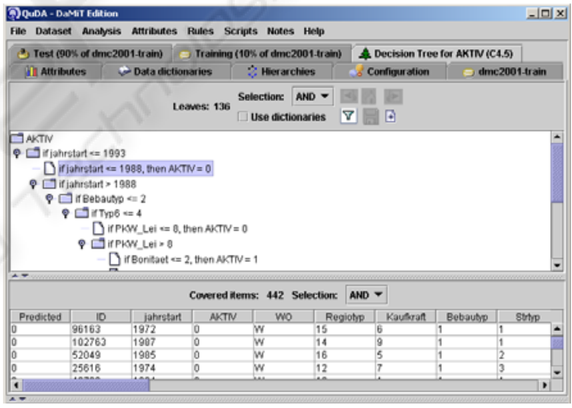

Figure 4: QuDA in doing Decision Tree Induction

This figure is displaying the generation and inspection of a decision tree by means of QuDA over the data of a realistic direct marketing case study. There is a node of the decision tree highlighted and all customers classified by this node are listed in the window below.

A user should check whether he agrees with an approximate classification like this. If not, he has to return to the tree generation process.

Data mining tools offer different ways to take a subtree of the classifier generated so far and continue model generation at the point picked up.

Following the exemplified procedure in an enterprise when dealing with the own problem on the own data is not only the right way to solve a problem, it is also an instance of learning when doing.

Recall that data mining is both an art and a science, and whatever we generate by means of data mining technologies and tools, we only do arrive at hypotheses. There can not be any guarantee at all that models (like, e.g., decision trees) generated over given data behave as successful as expected over other data in the future.

There is a need to verify generated models. How to do that appropriately, which alternatives do exist, and what the results mean and how they are backed up by statistics, can be studied in DaMiT – another case of learning when doing.

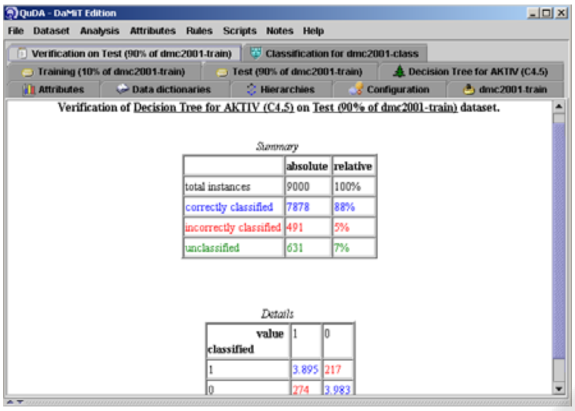

Figure 5: Verifying a Generated Decision Tree

The present Figure 5 shows a verification result for the decision tree of Figure 4 generated by means of the C4.5 implementation of QuDA.

Once a model has been established, as shown in Figure 6, it can be exported from the generation tool and saved for use in the future.

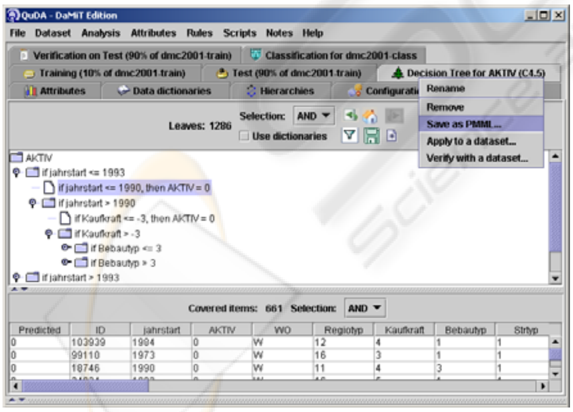

Figure 6: A Solution to the Application Problem

XML standards like PMML allow for an integration into an enterprise's IT infrastructure.

There is no closing sentence about learning when doing, because in areas like knowledge discovery and data mining, learning never ends. Systems like DaMiT are further developed to support this type of lifelong learning.

5 SUMMARY & CONCLUSIONS

On the one hand, the Internet in pervading the world has changed our daily life, and it is currently changing human learning at all stages ranging from schools through higher education to training in enterprises, research labs and governmental institutions and to lifelong learning. Technology enhanced learning is providing quite new opportunities.

On the other hand, there is the obviously eternal gap between academia and practice which appears as a certain reluctance to e-learning in practice.

Despite these obvious difficulties, we are at the cusp of closing the gap. As exemplified in the data mining domain, even academic e-learning has the urgent need of doing, i.e. **learning by doing**. In practice, for sure, there is the need of knowledgeable doing which leads to **learning when doing**. Enterprise application integration will allow for a proper dovetailing of learning and doing. Data mining may be an area worth for doing it now.

The present paper is based on work of colleagues and friends from 11 academic institutions and draws benefit from many other partners using the DaMiT system in their educational practice.

Note that as a courtesy to interested readers, there is an extended version of this paper available (see http://www.dfki.de/~jantke).

REFERENCES

Arikawa, S., Miyano, S., Shinohara, A., Kuhara, S., Mukouchi, Y., and Shinohara, T., 1993. A machine discovery from amino acid sequences by decision trees over regular patterns. *New Generation Computing*, 11, 361-375.

Grieser, G., Lange, S. and Memmel, M., 2003. DaMiT: Ein adaptives Tutorsystem für Data Mining. In *Von e-Learning bis e-Payment 2003*, K.P. Jantke, W.S. Wittig and J. Herrmann (eds.), infix, 192-203.

Grigoriev, P.A. and Yevtushenko, S.A., 2003. Elements of an agile discovery environment. In *6th Int. Conference on Discovery Science, DS 2003*. G. Grieser, Y. Tanaka and A. Yamamoto (eds.), Springer Verlag, LNAI 2843, 309-316.

Jantke, K.P., Memmel, M., Rostanin, O., Thalheim, B. and Tschiedel, B., 2003. Decision Support by Learning-On-Demand. In *CAiSE Workshop 2003*, Klagenfurt.

Linthicum, D., 2000. *Enterprise Application Integration*. Addison-Wesley.

Thalheim, B. and Düsterhöft, A., 2001. SiteLang: Conceptual Modeling of Internet sites. In *20th Int. Conference on Conceptual Modeling, ER'2001*. SpringerVerlag, LNCS 2224, 179-192.