Data Mining and Decision Support:  
A note on the issues of their integration and their relation to Expert Systems

Nada Lavrač
J. Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
Nada.Lavrac@ijs.si

Abstract. This paper provides some insights into the possible integrating aspects of two currently separate research areas Data mining (DM) and Decision Support (DS), and investigates the relation of DM and DS with Expert Systems (ES). In particular it investigates how DM can be used to support DS, and vice versa, how DS can be used to support DM. Finally, we investigate what future developments could lead to the integration of DM and DS.

1 Introduction

Data Mining (DM) [20, 7, 8, 22, 10] is concerned with finding patterns in data which are interesting (according to some user-defined measure of interestingness, e.g., with coverage above the requested threshold) and valid (according to some user defined measure of validity, e.g., classification accuracy). Numerous data mining algorithms exist, including the predictive data mining algorithms, which result in classifiers that can be used for prediction and classification, and descriptive data mining algorithm that serve other purposes like finding of associations, clusters, etc. The area has recently gained much attention of industry, due to the existence of large collections of data in different formats (including large data warehouses), and the increasing need of data analysis and comprehension. In addition to mining of data stored in data warehouses, e.g., in the form of relational data tables, there has recently been also increased interest in text and web mining.

Decision Support (DS) [13, 14] is concerned with developing systems aimed at helping decision makers solve problems and make decisions. Their main characteristics are that they incorporate both data and models; are designed to assist managers in semi-structured or unstructured decision-making processes; support, rather than replace, managerial judgment and are aimed at improving the effectiveness (rather than efficiency) of decisions. Decision Support Systems (DSS) can be data or model oriented. Data oriented DS tools (modern data warehouses, data cubes and OLAP, together with data visualization) involve no models, but enable good data understanding through segmentation, slicing, dicing, drilling-down, rolling-up and other operations. On the other hand, model oriented DS tools support the development of decision models in the form of
decision trees (notice that a decision tree as understood in decision support has a substantially different format from the decision trees used in data mining), influence diagrams and multi-attribute models.\(^1\)

In this paper we reflect upon the relation between data mining, decision support and expert systems. In particular we explore means of combining data mining with decision support, involving joint data preprocessing, standards for model exchange, and meta-learning providing decision support when choosing best data mining tools for a given problem.

2 Data mining, decision support and expert systems

Generally speaking, Decision Support Systems (DSS) are a broader notion than expert systems. There is a large intersection between the two, but some expert systems are not DSS.\(^2\) Expert systems that reproduce the reasoning process of a human decision maker can be categorized as process-oriented DSS.

A narrower definition of DSS taken in this paper is a system that helps humans to choose the best among the available alternatives. A DSS will evaluate all the alternatives (for example, all the individuals applying for a bank loan), ranking them in accordance with the system’s evaluation/utility function. On the other hand, when an expert system is used, for example, to diagnose patients, one is typically not interested in ranking patients according to some criterion, but rather in obtaining a prognostic/diagnostic outcome for an individual patient.

A typical expert system architecture consists of a knowledge base, an inference engine and a user interface, as shown in Figure 1. A DSS architecture proposal by E.G. Mallach [13], on the other hand, consists of a data base, a model base, possibly a knowledge base, and a user interface. Notice a clear distinction between the two: an expert system does not involve a database, whereas the correspondence between an expert system knowledge base, and a model and a knowledge base in a DSS, is not completely clear, since a DSS architecture should also explicitly include an analysis or inference engine. In comparison with a classical expert system architecture, a nowadays architecture of a DSS

\(^1\) The most straightforward and practical approach to developing a DSS seems to be by choosing one of the available DSS generators (shells) or data analysis tools, and the particular choice depends on the problem to be solved. Examples of good system generation shells include DATA (Decision Analysis) by TreeAge for decision trees, Analytica (A Software Tool for Uncertainty Analysis and Model Communication) by Lumina Decision Systems for influence diagrams, and DecisionPro by Vanguard Software Corporation for computational models. These systems support quantitative modeling and utility/probability propagation, whereas the DEX [2] shell (and its variant DEXi) support the development of qualitative multi-attribute utility models. Some attempts of combining several of the above mentioned representations have already been fruitful, such as the transformation of influence diagrams to a decision tree form, as supported by DATA.

\(^2\) For instance, if an expert system is used to automatically guide processes in an electric plant or a production line, it is not a DSS. On the other hand, if it is used to support managerial decisions, it certainly is a DSS as well.
could be as shown in Figure 2 [12], where the scheme is adapted for use in the field of medical decision making. The figure indicates that current DSS need to deal also with large volumes of data, as well as data gathering and analysis via the Internet and intranet.

The main source of knowledge for DM is data from which knowledge is extracted. On the other hand, in DS the main source is human knowledge, formalised in a format requested by a selected DS tool. There are many examples where the knowledge of experts and knowledge extracted from data have been integrated, so that the results of data mining feed into a decision support system, complementing expert knowledge. It is worthwhile noticing that a set of rules or a decision tree induced by a machine learning system can be viewed as an automatically constructed expert system knowledge base, and that the inference engine is simply rule firing or path finding in a decision tree, resulting in a probability distribution of outcomes (for instance, patient diagnoses). In this way, predictive machine learning can largely replace the standard expert system development methodology, provided that there is a sufficient amount of data (solved problems) available.

There are important approaches integrating expert provided knowledge and induced knowledge. The entire paradigm of multi-relational data mining [6], and inductive logic programming (ILP) in particular [17,11], employs background knowledge provided by experts as input to a learning system. In ILP, for instance, background knowledge is crucial for the success of learning. Another approach to the integration of expert knowledge and induced knowledge has been proposed in the development of an expert system for ECG diagnosis of cardiac arrhythmias [3]. The developed methodology proposes a semi-automatic ‘knowledge acquisition cycle’, involving a qualitative model construction, simulation of the model to construct an exhaustive database of examples, and inductive learning from examples to build a compact expert system knowledge base.

\footnote{In Figure 2, the box ‘temporal abstraction’ needs explanation. Data abstraction methods are used to glean out useful abstractions from raw numeric data, and temporal abstraction refers to data abstraction where the processed data is temporal.}
3 Integrating aspects of DM and DS

3.1 DS for DM

This section presents selected decision support methods, providing support in model selection (where models are developed by different DM algorithms or a single DM algorithm using different parameter settings), selection of the best algorithm for a given dataset, and model integration/combination. As outlined below, cost-sensitive classification supported by the ROC methodology can be used to find an optimal solution. Moreover, meta-learning can be applied to build rules or decision trees proposing the best classifier for a given classification task. Other meta-learning approaches to combine classifiers can be used.

- The well-known ROC (Receiver Operating Characteristic) methodology [18], initially used in medicine for cost-sensitive decision making, allows model selection using the ROC convex hull, indicating a tradeoff between specificity
and sensitivity of classifiers. A ROC curve indicates a tradeoff that one can achieve between the false alarm rate (1 - Specificity, plotted on the X-axis) that needs to be minimized, and the detection rate (Sensitivity, plotted on the Y-axis) that needs to be maximized.

Improved performance in terms of sensitivity, specificity and classification accuracy can be achieved by the adaptation of selected data mining methods to dealing with misclassification costs in the estimation of probabilities. More importantly, an appropriate sensitivity-specificity tradeoff, determined by the expert, can be achieved by applying different algorithms, as well as by different parameter settings of a selected data mining algorithm. In the context of developing a method for decision support for DM, the ROC method allows, through the construction of a convex hull of a set of points (results of classifiers), to identify and select models/classifiers that are optimal for a given sensitivity-specificity tradeoff. ROC analysis thus provides an integrated set of solutions, together with their optimality conditions in terms of sensitivity and specificity. Consequently, ROC analysis can be viewed as a decision support method for model selection and combination.

With increasingly many DM techniques to choose from, meta-learning seems to provide the means for successful industrial/commercial take-up of the DM technology. Meta-learning, e.g., as investigated in the EU funded project METAL (see http://www.metal-kdd.org/), aims at the development of methods and tools for providing support in model selection and method combination, by (a) selecting the best/most suitable model/algorithm to use on a given application, and (b) combining or integrating this with useful and effective transformations of the data. Automatic guidance in model selection and data transformation requires meta-knowledge. The use of inductive (meta-)learning techniques offer an automatic way of inducing meta-knowledge from experience as well as revising prior meta-knowledge, retrieved as cumulative expertise gained from ML/DM research and the conclusions of past comparative studies. Successful approaches to decision support in model selection and combinations have been developed, an example being Meta Decision Trees (MDT) [19] which provide a method for combining multiple classifiers. Instead of giving a prediction, MDT leaves specify which classifier should be used to obtain a prediction for a given dataset.

A less explicit and mostly automated decision support for ML is provided also by various approaches to combining of multiple classifiers (see [5] for

---

4 Sensitivity measures the fraction of positive cases that are classified as positive \( \frac{TP}{TP+FP} \), whereas specificity measures the fraction of negative cases classified as negative \( \frac{TN}{TN+FP} \). Sensitivity can be viewed as a detection rate that one wants to maximize. If the goal is to increase the sensitivity of answers, the learner should try to increase the correct classifications of positive cases (TP) and/or decrease the number of incorrect classifications of positive cases into class negative (FN). On the other hand, in order to increase the specificity, the learner should try to increase the number of correct classifications of negative cases (TN) and/or decrease the number of incorrect classifications of negative cases into class positive (FP). Note that 1 - Specificity can be interpreted as a false alarm rate which one wants to minimize.
an excellent overview), such as multistrategy learning, bagging, boosting, and other approaches to building optimal classifiers for a given dataset. Multistrategy learning [15] varies factors such as the choice of the learning method or the abstraction level to produce a variety of classifiers. In bagging [4] the learner remains the same but variety is achieved by learning from bootstrapped samples from the data; the final classification of obtained by majority voting. In boosting [9], on the other hand, the final classification is obtained after multiple classification runs by a single learner; in the next run, the data from which a classifier is learned is biased towards the training instances that were incorrectly classified in a previous classification run.

3.2 DM for DS

DM methods can be used to support the development of DS models. Two examples of using DM for DS are outlined.

- A multi-attribute decision support system DEX [2] incorporates a data mining component in the knowledge acquisition phase to semi-automatically build the hierarchical tree decomposing a decision problem into subproblems. This is done by asking the expert to evaluate a pre-defined set of partial decision making situations, and inducing or extrapolating a completed evaluation (utility) function from example solutions.

- The strongest link between data mining and decision support has been achieved by the system HINT [23] which enables the development of a DEX decision support model from data. HINT uses a function decomposition approach to develop a hierarchical decomposition of the decision problem into subproblems, thus automating part of the decision making process. For a given dataset, the constructive induction system HINT namely outputs a concept hierarchy, which needs to be developed manually in the standard DEX DS methodology. It was shown in a dozen of DS problems that HINT can indeed reconstruct experts’ decision knowledge.

3.3 Current trends in DM and DS integration

Despite this new commercial development, many integrating issues remain to be solved, and DM and DS integration techniques proposed. In addition to the integration issues of DM and DS described in two previous sections, the Sol-Eu-Net project (see http://soleunet.ij.s.si) is exploring further means of combining data mining with decision support, involving joint data preprocessing, standards for model exchange, and meta-learning providing decision support when choosing best data mining tools for a given problem. Some of the project partners (Czech Technical University, Dialogis GmbH, Bristol University, Oxford University and J. Stefan Institute) have been involved in the development of the following tools, that can be used in integrating data mining and decision support:
A pre-processing tool, based on the Sumatra scripting language [1], applicable for data pre-processing for both data mining and decision support. It allows access to various data sources, enabling simple definition of transformation tasks using a library of templates.

A common representation language supporting the exchange of data mining and decision support models for different application and visualisation tools. This development is built as an addition to the currently developing PMML (Predictive Model Markup Language) standard (see www.dmg.org). Its advantage is its independence of a selected application, platform and operating system.

Meta-learning tools for classifier selection, and ROC methodology for model selection (see the previous section).

Shared ontology, using the developing Sol-Eu-Net On Line Glossary of Terms SOGOT.

The RAMSYS methodology for solving data mining and decision support problems, requiring remote collaboration of project partners [21].

4 Conclusion

The two areas, data mining and decision support, are complementary, and there is also a strong link with expert systems. The most pressing integrating issue, however, is the integration of the database, data mining and decision support technology. Hopefully, OLE DB for Data Mining [16], to appear as part of the Microsoft SQL Server under the name Analysis Services, will provide a platform enabling cost and performance effective integration of data mining, decision support and information systems. Despite this new commercial development, many integrating issues remain to be solved, and integration techniques proposed in future work.

References