

# ON THE USE OF INFORMATION SYSTEMS RESEARCH METHODS IN DATA MINING

Mykola Pechenizkiy<sup>\*</sup>, Seppo Puuronen<sup>\*</sup> and Alexey Tsymbal<sup>\*\*</sup>

## 1. INTRODUCTION

Information systems are powerful instruments for organizational problem solving through formal information processing (Lyytinen, 1987). Data mining (DM) and knowledge discovery are intelligent tools that help to accumulate and process data and make use of it (Fayyad, 1996). Data mining bridges many technical areas, including databases, statistics, machine learning, and human-computer interaction. The set of data mining processes used to extract and verify patterns in data is the core of the knowledge discovery process. Numerous data mining techniques have recently been developed to extract knowledge from large databases.

The area of data mining is historically more related to AI (Artificial Intelligence), pattern recognition, statistical, and database communities, though we think there is no objective reason for that. And nowadays, although the field of data mining according to the ACM classification system<sup>\*\*\*</sup> for the computing field is a subject of database applications (H.2.8) that in sequence related to database management (H.2) and to information systems field (H.), there exists a gap between the data mining and information systems communities. Each of the two scientific communities publishes its own journals and books, and organizes different conferences that rarely cover the same issues. This situation is not beneficial since both communities share in common many similar problems being solved and therefore are potentially helpful for each other.

In this paper (in Section 2) we consider some existing frameworks for data mining, including database perspective and inductive databases approach, the reductionist statistical and probabilistic approaches, data compression approach, and constructive induction approach. We consider their advantages and limitations analyzing what these approaches account in the data mining research and what they do not.

The study of research methods in information systems by Järvinen (1999) encouraged us to analyse connections and appropriateness of them to the area of data mining. In Section 3 we are trying to view the data mining research as a continuous

---

<sup>\*</sup> Department of Computer Science and Information Systems, University of Jyväskylä, Jyväskylä, Finland

<sup>\*\*</sup> Department of Computer Science, Trinity College Dublin, Dublin, Ireland

<sup>\*\*\*</sup> See <http://www.acm.org/class/1998/ccs98.html>

information system development process. We refer to the traditional framework presented by Ives et al. (1980) that is widely known and has been used in the classification of Information Systems research literature. The framework is a synthesis of many other frameworks considered before by other researchers and covers their main elements. For us this framework is more substantial than the others since it also focuses on the development of information systems.

Ives et al. (1980) considers an information system (IS) in an organizational environment that is further surrounded by an external environment. According to the framework an information system itself includes three environments: user environment, IS development environment, and IS operations environment. Drawing an analogy to this framework we consider a data mining system as a special kind of adaptive information system that processes data and helps to make use of it. Adaptation in this context is important because of the fact that the data mining system is often aimed to produce solutions to various real-world problems, and not to a single problem. On the one hand, a data mining system is equipped with a number of techniques to be applied for a problem at hand. From the other hand there exist a number of different problems and current research has shown that no single technique can dominate some other technique over all possible data-mining problems (Wolpert and MacReady, 1996). Nevertheless, many empirical studies report that a technique or a group of techniques can perform significantly better than any other technique on a certain data-mining problem or a group of problems (Kiang, 2003). Therefore viewing data mining research as a continuous and never-ending development process of a DM system towards the efficient utilization of available DM techniques for solving a current problem impacted by the dynamically changing environment is a well-motivated position.

In this paper we focus on the IS development process. We consider information systems development framework of Nunamaker (1990-91) adapted to data-mining systems development. We discuss three basic groups of information systems research methods. Namely, we consider theoretical, constructive and experimental approaches with regard to Nunamaker's framework in the context of data mining. We demonstrate how these approaches can be applied iteratively and/or in parallel for the development of an artefact – a data-mining tool, and contribute to theory creation and theory testing. We conclude with a brief summary and discussion of our further research in Section 4.

## **2. THEORETICAL FRAMEWORKS FOR DATA MINING**

### **2.1. A database perspective and inductive databases**

A database perspective on data mining and knowledge discovery was introduced in Imielinski and Mannila (1996). The main postulate of their approach is: “there is no such thing as discovery, it is all in the power of the query language”. That is, one can benefit from viewing common data mining tasks not as dynamic operations constructing new pieces of information, but as operations finding unknown (i.e. not found so far) but existing parts of knowledge.

In Boulicaut et al. (1999) an inductive databases framework for the data mining and knowledge discovery in databases (KDD) modeling was introduced. The basic idea here is that data-mining task can be formulated as locating interesting sentences from a given logic that are true in the database. Then discovering knowledge from data can be

viewed as querying the set of interesting sentences. Therefore the term “an inductive database” refers to such a type of databases that contains not only the data but a theory about the data as well (Boulicaut et al., 1999).

This approach has some logical connection to the idea of deductive databases, which contain normal database content and additionally a set of rules for deriving new facts from the facts already present in the database. This is a common inner data representation. For a database user, all the facts derivable from the rules are presented, as they would have been actually stored there. In a similar way, there is no need to have all the rules that are true about the data stored in an inductive database. However, a user may imagine that all these rules are there, although in reality, the rules are constructed on demand. The description of an inductive database consists of a normal relational database structure with an additional structure for performing generalizations. It is possible to design a query language that works on inductive databases (Boulicaut et al., 1998). Usually, the result of a query on an inductive database is an inductive database as well. Certainly, there might be a need to find a solution about what should be presented to a user and when to stop the recursive rule generation while querying. We refer an interested reader to the work of Boulicaut et al. (1999).

## 2.2. The reductionist approach

In Mannila (2000) two simple approaches to the theory of data mining are analysed. The first one is the reductionist approach of viewing data mining as statistics. Generally, it is possible to consider the task of data mining from the statistical point of view, emphasizing the fact that DM techniques are applied to larger datasets than it is in statistics. And in this situation the analysis of the appropriate statistics literature, where strong analytical background is accumulated, would solve most of the data mining problems. Many data mining tasks naturally may be formulated in statistical terms, and many statistical contributions may be used in data mining in a quite straightforward manner. The second approach discussed by Mannila (2000) is a probabilistic approach. Generally, many data mining tasks can be seen as the task of finding the underlying joint distribution of the variables in the data. Good examples of this approach would be Bayesian network or a hierarchical Bayesian model, which give a short and understandable representation of the joint distribution. Data mining tasks dealing with clustering and/or classification fit easily into this approach. However, it should be admitted that data mining researchers with computer science background typically have rather little education in statistics and this is a reason to the fact that achievements from statistics are used not to such an extent as could be possible.

A deeper consideration of data mining and statistics shows that the volume of the data being analysed and different background of researchers are, probably, not the most important ones that make the difference between the areas. Data mining is an applied area of science and limitations in available computational resources is a big issue when applying results from statistics to data mining. The other important issue is that data mining approaches emphasize database integration, simplicity of use, and understandability of results. Last but not least Mannila (2000) points out that the theoretical framework of statistics does not concern much about data analysis as a process that generally includes data understanding, data preparation, data exploration, results evaluation, and visualisation steps. However, there are persons (mainly with

strong statistical background) who equate DM to applied statistics, because many tasks of DM may be perfectly represented in terms of statistics.

### 2.3. Data compression approach

A data compression approach to data mining can be stated in the following way: compress the dataset by finding some structure or knowledge for it, where knowledge is interpreted as a representation that allows coding the data by using fewer amount of bits. For example, the minimum description length (MDL) principle (Mehta et al., 1995) can be used to select among different encodings accounting to both the complexity of a model and its predictive accuracy.

Machine learning practitioners have used the MDL principle in different interpretations to recommend that even when a hypothesis is not the most empirically successful among those available, it may be the one to be chosen if it is simple enough. The idea is in trading between consistency with training examples and empirical adequacy by predictive success as it is, for example, with accurate decision tree construction. Bensusan (2000) connects this to another methodological issue, namely that theories should not be ad hoc that is they should not overfit the examples used to build them. Simplicity is the remedy for being ad hoc both in the recommendations of philosophy of science and in the practice of machine learning.

The data compression approach has also connection with the rather old Occam's razor principle that was introduced in 14th century. The most commonly used formulation of this principle in data mining is "when you have two competing models which make exactly the same predictions, the one that is simpler is the better".

Many (if not every) data mining techniques can be viewed in terms of the data compression approach. For example, association rules and pruned decision trees can be viewed as ways of providing compression of parts of the data. Clustering approaches can also be considered as a way of compressing the dataset. There is a connection to Bayesian theory for modelling the joint distribution – any compression scheme can be viewed as providing a distribution on the set of possible instances of the data.

However, in order to produce a structure that would be comprehensible to the user, it is necessary to select such compression method(s) that is (are) based on concepts that are easy to understand.

### 2.4. Constructive induction approach

Constructive induction is a learning process that consists of two intertwined phases, one of which is responsible for the construction of the "best" representation space and the second concerns with generating hypothesis in the found space (Michalski and Wnek, 1993). Constructive induction methods are classified into three categories: data-driven (information from the training examples is used), hypothesis-driven (information from the analysis of the form of intermediate hypothesis is used) and knowledge-driven (domain knowledge provided by experts is used) methods. Any kind of induction strategy (implying induction, abduction, analogies and other forms of non-truth preserving and non-monotonic inferences) can be potentially used. However, the focus usually is on operating higher-level data-concepts and theoretical terms rather than pure data. Michalski (1997) considers constructive (expands the representation space by attribute

generation) and destructive (contract the representational space by feature selection or feature abstraction) operators that can be applied to produce a better representation space comparing to the original one. In Bensusan (1999) it was shown that too many theoretical terms could impair induction. This vindicates an old advise of the philosophy of science: avoid adding unnecessary metaphysical baggage to a theory. Theoretical terms are often contrasted with observational terms. It is generally accepted that the more data we have the better model we can construct. However, this is not true for higher-level concepts that constitute a theory.

Many data mining techniques that apply wrapper/filter approaches to combine feature selection, feature extraction or feature construction processes (as means of dimensionality reduction and/or as means of search for better representation of the problem) and a classifier or other type of learning algorithm can be considered as constructive induction approaches.

## **2.5. Conclusion on considered frameworks**

The reductionist approach of viewing data mining in terms of statistics has advantages of the strong theoretical background and easy-formulated problems. The data compression and constructive induction approaches have relatively strong analytical background, as well as connections to the philosophy of science. In addition to the just-mentioned frameworks an interesting solution is proposed in the microeconomic view on data mining, introduced by Kleinberg (1998), where a utility function is constructed and trying to be maximized. The data mining tasks concerning processes like clustering, regression, and classification fit easily into these approaches.

The inductive databases framework suggests architecture for data mining systems and allow to view data mining as a process. Association rules and other simple pattern formalisms can be described by this approach. However, for example, clustering is harder to describe in a useful way within the inductive databases framework.

In one way or another, we can easily see the exploratory nature of the frameworks for the data-mining field. Different frameworks account different data mining tasks, allow preserving and presenting background knowledge. However, what seems to be lacking in most of the approaches, are the ways for taking the iterative and interactive nature of the data mining process into account (Mannila, 2000). Furthermore, none of the considered frameworks considers data mining in the context of an adaptive system that processes information.

In the next section we introduce an information systems development framework and then consider how data mining can be seen as an iterative and interactive development process within this framework.

## **3. DATA MINING AND INFORMATION SYSTEMS FRAMEWORK**

### **3.1. Generations of DM systems**

Present history of data mining systems' development totals three main stages/generations (Piatetsky-Shapiro, 2000). Year 1989 can be referred to as the first generation of data mining/KDD systems when a few single-task data mining tools such as C4.5 decision tree algorithm (Quinlan, 1993) existed. They were difficult to use and

required significant preparation. Most of such systems were based on a loosely-coupled architecture, where the database and the data mining subsystems were realised as separate independent parts. This architecture demands continuous context switching between the data-mining engine and the database (Imielinski and Mannila, 1996).

Then, the year 1995 can be associated with formation of the second-generation tools-suits. Data mining as a core part of KDD started to be seen as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad, 1996, 22). Some examples of the knowledge discovery systems that follow Fayyad’s view on DM as the process are: SPSS Clementine, SGI Mineset (Brunk et al., 1997), and IBM Intelligent Miner (Tkach, 1998).

Numerous KDD systems have recently been developed. At the beginning of the millennium there exist about 200 tools that could perform several tasks (clustering, classification, visualization) each for specialized applications (therefore often called “vertical solutions”) (Piatetsky-Shapiro, 2000). This growing trend towards integrating data-mining tools with specialized applications has been associated with the third generation of DM systems (Fayyad and Uthurusamy, 2002).

Because of increasing number of such “vertical solutions” and possibility to accumulate knowledge from these solutions, there is a growing potential for appearance of next-generation database mining systems to manage KDD applications. These systems should be able to discover knowledge by selecting and combining several available most suitable for specific domain KDD techniques. While today’s algorithms tend to be fully automatic and therefore fail to allow guidance from knowledgeable users at the key stages in the search for data regularities, the researchers and the developers, who are involved into the creation of the next generation data mining tools, are motivated to provide a broader range of automated steps in the data mining process and make this process more mixed-initiative, in which human experts collaborate more closely with the computer to form hypotheses and test them against the data (Ankerst, 2002).

Since a data mining system is often aimed to produce solutions not to a single problem but rather to various real-world problems, it has to be armed with a number of techniques to be applied for a problem at hand. However, current research has shown that no single technique can dominate some other technique over all possible data-mining problems (Wolpert and MacReady, 1996). Nevertheless, many empirical studies report that a technique or a group of techniques can perform significantly better than any other technique on a certain data-mining problem or a group of problems (Kiang, 2003). Therefore a good data mining system should be adaptive for solving a current problem impacted by the dynamically changing environment and being continuously developed towards the efficient utilization of available DM techniques.

### 3.2. Information systems framework

The traditional framework presented by Ives et al. (1980) is widely known and has been used in the classification of IS research literature. We consider this framework because:

- (1) it is a synthesis of many other frameworks considered before by other researchers and covers their main elements;
- (2) it is helpful in drawing the analogy between the information systems and data mining systems as a special kind of adaptive information system that processes data and helps to make use of it;

- (3) for us this framework is more substantial than the others since it also focuses on the development of information systems as we focus on the development of data mining systems.

Ives et al. (1980) considers an information system in an organizational environment that is further surrounded by an external environment. According to the framework an information system itself includes three environments: user environment, IS development environment, and IS operations environment. There are accordingly three processes through which an IS has interaction with its environments: the use process, the development process, and the operation process.

Analogously, a data-mining system that is equipped with a collection of DM techniques and knowledge how to utilize those for various tasks can be considered as a system with user environment, DM development environment, and DM operations environment. However, in this paper, we focus on the development process of an artefact for data mining and leave operation and use processes for further research.

In the information systems research a variety of research methods have been applied. Davis (2000, 80) expresses this saying that “the field has a richer set of views than other fields because the positivist philosophy that dominated the American research and the phenomenology philosophy that tended to dominate in Europe were both supported by the worldwide community”. Even when there are still discussions going on about suitable research methods in the field, we share with many others the opinion that there is room for many research methods, both hard and soft.

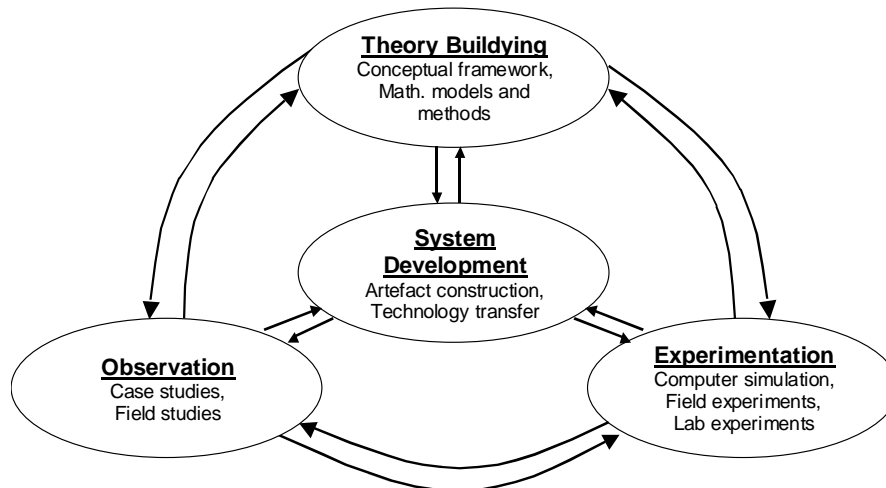
Iivari et al. (1998) relate development process to the constructive type of research because of their philosophical belief that development always involves creation of some new artefacts – conceptual (models, frameworks) or more technical artefacts (software implementations). The research approach is classified as constructive where scientific knowledge is used to produce either useful systems or methods, including development of prototypes and processes. Iivari et al. (1998) argue the importance of constructive research especially for applied disciplines of information systems and computer science, and DM may be considered as such a discipline.

Nunamaker et al. (1990-91, 94) consider system development as a central part of a multi-methodological information systems research cycle (Figure 1). Theory building involves discovery of new knowledge in the field of study, however it is rarely contributing directly to practice. Nevertheless, the built theory often (if not always) needs to be tested in the real world to show its validity, recognize limitations and make refinements according to observations made during its application. Therefore research methods are subdivided into basic and applied research, as naturally both are common for any large project (Nunamaker et al., 1990-91). A proposed theory leads to the development of a prototype system in order to illustrate the theoretical framework from the one hand, and to test it through experimentation and observation with subsequent refinement of the theory and the prototype in an iterative manner. Such a view presents the framework of IS as a complete, comprehensive and dynamic research process. It allows multiple perspectives and flexible choices of methods to be applied during different stages of the research process.

In the following subsections we consider applying information systems research methods in the context of the data-mining field. we consider theoretical, constructive and experimental approaches with regard to Nunamaker’s framework in the context of data mining. We demonstrate how these approaches can be applied iteratively and/or in

parallel for the development of an artefact – a data-mining tool, and contribute to theory creation and theory testing.

Particularly, in the next section we consider the construction of an artefact for data mining as system development applying multi-methodological information systems research cycle presented in this section



**Figure 1.** A multimethodological approach to the construction of an artefact for data mining (adapted from Nunamaker et al., 1990-91, 94).

### 3.3. Construction of an artefact for data mining

Can we build an artefact that would be useful? If a research question deals with the verbs like introduce, improve, maintain, cease, extend, correct, adjust, enhance and so on, the study according to Järvinen (1999, 59) likely belongs to the constructive research. Indeed these are the actions that researchers in the area of data mining perform when developing new theories and their applications.

From the data mining research point of view the constructive approach can be seen to help to manipulate and coordinate integrative work (selection and combination) of different data mining techniques, and to conduct the experimental approach. However, in this paper we emphasize the goal of a data mining artefact construction as the major one.

Development of an artefact for data mining can be described in terms of initial and target/final states and the building process itself that includes specification and implementation stages (and usually it is difficult to see if these stages are performed sequentially, iteratively or in parallel). The building process aims to minimize the difference between the target and final states. In our situation the initial state may be described in terms of existing (available in the system) sets of different data mining techniques, e.g. certain clustering, feature transformation and classification techniques. And the target state would be a system that has a possibility adaptively select/construct the most appropriate approach/solution for a given task according to the specificity of this given task.



It is obvious that in order to construct a good artefact with such adaptivity we need some background knowledge about the artefact's components (that is basic data mining techniques) and their appropriateness for certain dataset characteristics. Beside this we need also some background knowledge about the artefact's external environment that are different real-world problems, often called just datasets.

In data mining a dataset is usually characterised by analysis of its domain, statistical, information-theoretical properties and simple measures like the number of instances and attributes. And DM techniques are commonly specified with their requirements, capabilities and/or limitations. Examples of such characteristics are algorithm run-time parameters, ability of handling misclassification costs, and data types supported. Beside specifications, DM techniques may be characterised by many representational and functional characteristics associated either with the expert knowledge about the techniques or with the past learning experience of a corresponding DM technique. These various characteristics include attribute types supported, bias/variance profile, incrementality, cost handling support; efficiency characteristics: training and execution time, training and execution space; resilience characteristics: scalability, tolerance to missing values, tolerance to noise and irrelevant and redundant attributes; and finally practicality characteristics: runtime parameter handling, interpretability, and transparency (Hilario and Kalousis, 1999).

Thus, it is natural that the theory-creating research has to be performed during which the basics of available data mining techniques should be elaborated. For this purpose a literature survey and review commonly are undertaken. This helps to understand the background of the problem and analyse previous findings in the area. However, such theory-creating research can be supported also by meta-learning approaches that in (semi)-automatic way may help to state and check different hypothesis about the relations between technique's and dataset's characteristics. An overview of different meta-learning approaches can be found, for example, in Hilario and Kalousis (2000).

From the theory development point of view there are possibilities to apply either inductive or deductive approaches, and actually it is reasonable to try their combination in the sense that it is possible to use the findings from both approaches in order to check their consistency and guarantee more sophisticated completeness. Inductive theory building is based on search for trends, generalizations from experiments, whereas deductive approaches are based on logical inference on a set of axioms/hypothesizes.

It should be noticed that in some cases it is not possible just to adjust an existing program for someone's research purposes and program design and implementation are required. However it is reasonable to use existing libraries and appropriate tools when possible. In this situation it might be possible to use tested and validated tools as a core/backbone for a new tool and the development process can be focused on the new part of the desired tool.

There are two alternatives to create a tool: to develop it in whole or to develop one component of the tool after another. The second alternative has the following advantages: each component can be designed, implemented, tested, and refined independently before it is included into the meta-approach. The control over the individual components can be organized and the experiments can be easily performed on separate components also.

Evaluation process is an essential part of constructive research. Usually, experimental approach is used to evaluate a data mining artefact. We consider the

experimental approach in the next subsection. We will try to show that the experimental approach, however, can be beneficial for theory testing and can be a means of constructing new pieces of knowledge and thus contributing to the theory-creating process.

### **3.4. Experimental approach: theory testing and artefact evaluation**

By the evaluation of artefact we understand first of all (1) the evaluation of learned models and meta-level models and (2) testing the constructed theory of different data processing and machine-learning techniques selection and combination.

As from the theory evaluation as from the artefact evaluation point of view, the general principle of evaluation – the new derivation or construct must be better than its best challenger – is applicable for data mining as well. ‘Goodness’ criterion of a built theory or an artefact is multidimensional and sometimes is difficult to be defined because of mutual dependence between the compromising variables. However, it is more or less easy to construct a criterion based on such estimates as accuracy (including sensitivity and specificity, and various costs matrices) of a built model and its performance (time and memory resources). On the other hand – it is more difficult or even impossible to include into a criterion such important aspects as interpretability of the artefact’s output because such kinds of estimate usually are subjective and can be evaluated only by the end-users of a system.

Experimental study can be done in the ‘field’ or in the ‘laboratory’. In the first case different approaches are tested on so-called real-world datasets with real users. In the second case systematically controlled experiments can be organized. Controlled experiments sometimes might produce more beneficial results for theory creating, since unlike real world datasets, synthetically generated data allow to test exactly the desired number of characteristics while keeping all the others unchanged.

Theory testing might be seen here at different levels. A low-level task is to evaluate how well a built model works. The other task is to analyse how the built model performs comparing to the other models. Then it is usually necessary to compare the algorithm selected to build the models with other algorithm(s). Finally, when ‘laboratory’ experiments and evaluation are finished (that are experiments on synthetic datasets in our situation), it is necessary to go to the field and organize ‘field’ experiments (that would be experiments on real-world or benchmark datasets).

When testing and validating a model, data miners use several techniques. They include sampling, validation, cross-validation, stratification, Monte Carlo methods, division a dataset into training, validating and testing sets etc. There are two of the most essential elements of any experimental design, namely randomization and experimental control (of all nuisance variables or it is better to say possibility to control adjustable variables and restrictions of known factors).

The evaluation of a selected approach can be provided either based on the filter paradigm, when evaluation process is independent from a learning algorithm and the most appropriate approach is chosen from available ones according to certain data characteristics before the algorithm starts, or based on the wrapper paradigm (Kohavi, 1998) that assumes interaction between the approach selection process and the performance of the integrative model. In order to compare the two approaches Student’s t-test and McNemar’s test are used as standard de facto (Dietterich, 1998).

However, the experimental approach benefits not only for the artefact evaluation and theory testing that has been used for artefact construction but also it can contribute to the knowledge by producing new pieces of theory about selection and/or combination of DM techniques for a given dataset. Meta-learning approaches is one good example of such attempts to contribute to new pieces of theory induction.

In conclusion we would like to notice that it is reasonable to consider how the results achieved through different research approaches relate to each other and search for contradictions in the results. It can be expected that such joint use of these approaches will give a better understanding of the introduced research goal and benefit in a more significant and sophisticated contribution to the knowledge in the area.

#### 4. CONCLUSION

In this paper we considered several frameworks for data mining. These frameworks are based on different approaches, including inductive databases approach, the reductionist statistical approaches, data compression approach, constructive induction approach and some others. We considered advantages and limitations of these frameworks. We presented the view on data mining research as continuous and never-ending development process of an adaptive DM system towards the efficient utilization of available DM techniques for solving a current problem impacted by the dynamically changing environment. We discussed one of the traditional information systems frameworks and, drawing the analogy to this framework, we considered a data mining system as the special kind of adaptive information system. We adapted the information systems development framework for the context of data-mining systems development. Three basic groups of information systems research methods, applicable for data mining research were discussed, including theoretical, constructive, and experimental approaches. We demonstrated how these approaches could be applied iteratively for the development of a data-mining system. The theoretical backgrounds need to be exploited during the constructive work and the constructed artefact can be used for experimentation. The results of constructive and experimental work can be used to refine theory. Thus, all the research approaches are heavily connected to each other.

In this paper we considered the development process of a data mining system and the constructive research as main means that needs to be supported by the theory-testing research. Further analysis of IS framework briefly presented in subsection 3.2 can be beneficial in the context of the data mining artefact construction and use.

Beside traditional IS framework considered for data mining, adaptation of a knowledge management framework and knowledge engineering perspective towards data mining framework construction is the other direction of our further research.

#### Acknowledgments

This research is partly supported by the COMAS Graduate School of the University of Jyväskylä, Finland and by the Science Foundation Ireland under Grant No. S.F.I.-02IN.11111.

## REFERENCES

- Ankerst, M. 2002, Report on the SIGKDD-2002 panel the perfect data mining tool: interactive or automated. *SIGKDD Explorations* **4**(2), 110-111.
- Bensusan, H. 1999, Automatic bias learning: an inquiry into the inductive basis of induction. PhD thesis, School of Cognitive and Computing Sciences, University of Sussex.
- Boulicaut, J., Klemettinen, M., and Mannila, H. 1998, Querying inductive databases: a case study on the MINE RULE operator. In *Proceedings of 2nd European Symposium on Principles of Data Mining and Knowledge Discovery* (PKDD'98), France, pp. 194-202.
- Boulicaut, J., Klemettinen, M., and Mannila, H. 1999, Modeling KDD processes within the inductive database framework. In *Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery*, Springer-Verlag, London, UK, 293-302.
- Brunk, C., Kelly, J., and Kohavi, R. 1997, MineSet: an integrated system for data mining. In Heckerman D., Mannila H., Pregibon D. (Eds.) *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining* (KDD-97), AAAI Press, California, 135-138.
- Davis, G. 2000, Information systems conceptual foundations: looking backward and forward. In R. Baskeville, J. Stage, J. DeGross (Eds.), *Organizational and Social Perspectives on IT*, Proceedings of the IFIP International Working Conference on the Social and Organizational Perspective on Research and Practice in Information Technology. Boston: Kluwer, 61-82.
- Dietterich, T. 1998, Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, **10** (7), 1895-1923.
- Fayyad, U.M. 1996, Data mining and knowledge discovery: making sense out of data, *IEEE Expert* **11**(5), pp. 20-25
- Fayyad, U.M., and Uthurusamy, R. 2002, Evolving data into mining solutions for insights. *Communications of the ACM* **45**(8): 28-31
- Hilario, M., and Kalousis, A. 1999, Characterizing learning models and algorithms for classification. CUI – University of Geneva, TR UNIGE-AI-9-01.
- Hilario, M., and Kalousis, A. 2000, Building algorithm profiles for prior model selection in knowledge discovery systems. *Engineering Intelligent Systems, Special Issue on Data Mining*, **8**(2), IEEE Press.
- Iivari, J., Hirscheim, R., and Klein, H. 1998, A paradigmatic analysis contrasting information systems development approaches and methodologies, *Information Systems Research*, **9**(2), 164-193
- Imielinski, T., and Mannila, H. 1996, A database perspective on knowledge discovery. *Communications of the ACM*, **39**(11), 58-64.
- Ives, B., Hamilton, S., and Davis, G., 1980, A framework for research in computer-based management information systems. *Management Science*, **26**(9), 910-934.
- Järvinen, P. 1999, On research methods. Tampere: Opinpaja. (<http://www.uta.fi/~pj/>)
- Kiang, M. 2003, A comparative assessment of classification methods, *Decision Support Systems* **35**, 441-454.
- Kleinberg, J., Papadimitriou, C., and Raghavan, P. 1998, A microeconomic view of data mining, *Data Mining and Knowledge Discovery* **2**(4), 311-324
- Kohavi, R., and John, G. 1998, The wrapper approach. In: *Feature Selection for Knowledge Discovery and Data Mining*, H. Liu and H. Motoda (Eds.), Kluwer Academic Publishers, pp. 33-50.
- Lyytinen, K., 1987, Different perspectives on information systems: problems and solutions. *ACM Computing Surveys*, **19**(1), 5-46.
- Mannila, H. 2000, Theoretical framework for data mining. *SIGKDD Explorations* **1**(2).
- Mehta, M., Rissanen, J., and Agrawal, R. 1995, MDL-based decision tree pruning. In U.M. Fayyad, R. Uthurusamy (Eds.) *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (KDD 1995), AAAI Press, Montreal, Canada, 216-221.
- Michalski, R.S., and Wnek, J. 1993, Constructive induction an automated design of knowledge representation spaces for machine learning. Reports of the Machine Learning and Inference Laboratory, MLI 93-11, School of Information Technology and Engineering, George Mason University, Fairfax, VA, November.
- Michalski, R.S. 1997, Seeking knowledge in the deluge of facts, *Fundamenta Informaticae* **30**, pp. 283-297.
- Nunamaker, W., Chen, M., and Purdin, T. 1990-91, Systems development in information systems research, *Journal of Management Information Systems*, **7**(3), 89-106.
- Quinlan, J.R. 1993, C4.5 Programs for machine learning. San Mateo CA: Morgan Kaufmann.
- Piatetsky-Shapiro, G. 2000, Knowledge discovery in databases: 10 years after. *SIGKDD Explorations* **1**(2).
- Tkach, D. 1998, Information mining with the IBM intelligent miner family. An IBM Software Solutions White Paper. [www.acm.org/sigs/sigmod/disc/disc99/disc/ibm/whitefam3.pdf](http://www.acm.org/sigs/sigmod/disc/disc99/disc/ibm/whitefam3.pdf)
- Wolpert, D. H. and MacReady, W. G. 1996, No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, **1**(1), 67-82.