2000

Enhancing Government Decision Making through Knowledge Discovery from Data

Herna Viktor
University of Pretoria, hlviktor@hakuna.up.ac.za

Heidi Arndt
University of Pretoria

Mauritz Oberholzer
University of Pretoria

search

Enhancing Government Decision Making through Knowledge Discovery from Data

Herna Viktor, Heidi Arndt and Mauritz Oberholzer

Department of Informatics
School of Information Technology
University of Pretoria
Pretoria, 0002, South Africa
Email: hlviktor@hakuna.up.ac.za
Phone: +27 12 420 3376
Fax: +27 12 362 5287

***Abstract* - A major challenge facing management in developed countries is improving the performance of knowledge and service workers, i.e. the decision makers. In a developing country such as South Africa, with a well-developed business sector, the need to improve the performance of decision makers, especially in government, is even more crucial. South Africa has to face many new challenges in the 21st century - growing environmental concerns, massive social and economic inequalities, an ageing population, low productivity, massive unemployment and the nation's evolving role in Africa. The importance of science and technology to address these pressing issues cannot be overemphasised.**

This paper discussed the development of a knowledge-base to aid government decision makers in interpreting the results of the National Research and Technology (NRT) Audit that was undertaken by the South African Department of Arts, Culture, Science and Technology. An intelligent data analysis tool is employed to construct a knowledge-base, using a data-driven rather than a knowledge-driven approach to knowledge-base con-struction. The knowledge-base is constructed directly from the data as contained in the NRT Audit data warehouse. The rules contained in the knowledge-base are produced by a team of data mining techniques that cooperate as members of a learning system. This knowledge-base is used to augment the knowledge of the human experts. Results show that the information, as discovered during the knowledge-base construction process, either enhanced or contradicted the finding of the human experts.

I. Introduction

It has been estimated that the amount of information in the world doubles every 12 months [1]. Large data warehouses store terabytes of data that may contain information that is crucial to ensure effective management. Due to this explosion of data, many decision makers have realised the importance of systems that produce timely, relevant information in order to make informed decisions. Knowledge discovery from data (KDD) is a new field of research that addresses this need for intelligent data analysis. This approach provides a method for constructing a *knowledge-base* directly from domain data. The KDD process is structured into six phases, namely data selection, cleaning, enrichment, coding, discovery and reporting. The actual discovery stage is called data mining. A number of data mining techniques exist, including rule induction algorithms, decision trees and neural networks. Each of these techniques form descriptions of the concepts as contained in a set of training data, thus potentially providing new insights into the underlying data.

This paper discusses the construction of a *knowledge-base* for the South African Department of Arts, Culture, Science and Technology (DACST), using a KDD approach. This *knowledge-base* will be used to aid the governmental decision makers when formulating a new Knowledge and Technology Policy Framework for South Africa. The *knowledge-base* was constructed from the data, as contained in a NRT data warehouse, thus eliminating the time consuming process that is normally associated with knowledge acquisition. In this approach, data mining techniques work together to build a team *knowledge-base* which combines the best individual results.

The paper is organised as follows. Section 2 described the National Research and Technology (NRT) Audit on which this case study is based. The learning approach followed to construct the *knowledge-base* is introduced in Section 3. Section 4 discusses the results obtained when applying the KDD approach to a portion of the NRT data. Section 5 concluded the paper and highlights areas of further research.

II. National Research and Technology Audit

The role of information, communication and knowledge in shaping socio-economic development has become a top concern for African countries. In this regard South Africa is

no exception to the rule. The new government of South Africa has made it one of its central aims to eliminate social and economic inequities and to create new opportunities for the country's most disadvantaged population sectors. South African government departments have launched an aggressive series of initiatives which seek to achieve broadbased growth and equitable development through communications and information technologies. At the same time it is the overriding purpose of the government to establish policies and practical programmes that will improve the quality of life for all South Africans. One of these initiatives conducted by the South African DACST was a National Research and Technology Audit to determine the strengths and weaknesses of the current science and technology system. A major output of the Audit was a NRT Audit data warehouse. The data warehouse was created to provide the data needed for formulating a Knowledge and Technology Policy Framework for this country. This framework will be directed at increasing the effectiveness of technological innovation and will act as a contributor to increase South Africa's industrial productivity, environmental sustainability and international competitiveness. Another output of the Audit was the Knowledge Synthesis Report. This report produced a number of findings that described the current state of science and technology in South Africa and outlined certain trends. These findings were based on the opinions of a number of domain experts as well as the analysis of the information as contained in a number of questionnaires. The information contained in the Knowledge Synthesis Report is to be used for future decision making.

Analysis showed that the findings, as stated in the Report, did not address all of the questions as raised prior to the NRT Audit. Rather, it contained general statements concerning the current state of science and technology. The statements were not based on the detailed information as contained in the NRT Audit data warehouse. *For example, the data indicated that the technologies used by the footware and textiles industries were outdated, leading to the inability of these industries to compete internationally. The report did not include this inportant information.*

None of the domain experts had adequate knowledge of the current state-of-the-art in all of the disciplines and the individuals were therefore biased towards their areas of expertise. *The report did not, for example, include a detailed analysis of the Information Technology and Telecommunication industries.* Also, a large portion of useful information contained in the original questionnaires was lost due to the restrictive nature of the computerization process.

In an attempt to address these shortcomings, a KDD project was initiated. The primary aim of this project, which is the subject of this paper, was therefore to determine whether the data as contained in the NRT data warehouse supports the findings as contained in the Knowledge Synthesis Report. The second objective was to use the KDD approach to obtain new, interesting insights into the state of Research and Technology in South Africa. These insights can be used to test the aptness of the policies included in the final Knowledge and Technology Policy Framework.

III. Decision Support through Knowledge Discovery

Automated problem solving has been a target for generations, including the development of statistical models such as regression or forecasting, and management science models used for inventory level determination and the allocation of resources [7]. The development of new automated approaches to decision support, which attempt to find relevant information from huge amounts of data, is one of the main challenges of software developers today [1]. Data mining techniques address this need through the discovery of hidden knowledge, unexpected patterns and new rules from data sources. Diverse data mining techniques produce complementary results when supplied with exactly the same data set. This is due to the inherent preferences (the so-called inductive bias) that each technique uses during the discovery step [3]. The grouping of more than one heterogeneous technique into a cooperative learning system, in which the data mining techniques learn from one another, alleviates this shortcoming.

Machine learning refers to a branch of data mining that attempts to discover knowledge through the analyses of historical cases [1,7]. The section discusses a data mining approach that combines diverse machine learning techniques, together with the human experts, into a cooperative learning framework. By using several cooperating techniques, it is possible to discover a broader range of information, thereby constructing a *knowledge-base* representing the rules as hidden in the data.

A. Learning system

The cooperative inductive learning team (CILT) learning system produces sets of high quality concept descriptions, in the form of IF-THEN rules, to be used for classification purposes [8]. The system is heterogeneous and consists of two or more data mining techniques, henceforth referred to as machine learners, together with zero or more human learners that cooperate to learn, as depicted in Figure 1.

The figure shows two communicating learners that each consists of a learning component and a *knowledge-base*, which stores the learner's current knowledge in the form of rules and quality measures.

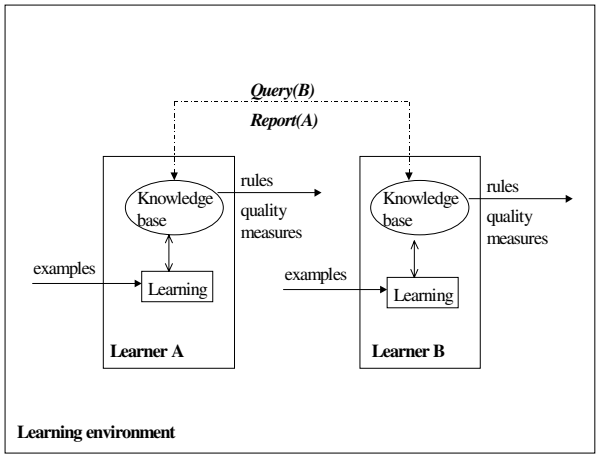

Fig. 1: Learner A queries Learner B's *knowledge-base*

Cooperative learning has two objectives. Firstly, each individual learner aims to produce an individual set of high quality rules. Secondly, a team *knowledge-base* that contains the combined results is created. The learning process consists of two stages, as discussed next.

The first stage involves individual learning. Each machine learner executes an algorithm to produce an initial set of rules. This set of rules is placed in the machine learner's *knowledge-base*. A human subject's knowledge, on the other hand, is used to form a set of rules that is placed in the human learner's *knowledge-base*. This process is facilitated by a knowledge engineer who acquires the knowledge from the human subject.

The second stage concerns cooperative learning. Here, the learners cooperate by accessing one another's knowledge by means of a querying/reporting tool. A learner learns by using the knowledge obtained from the other learners to augment its *knowledge-base*. All team members cooperate to solve the learning problem at hand and the whole team is required to participate in cooperative learning. The results of cooperative learning are placed in a team *knowledge-base* that reflects the team effort.

B. Learner architecture

A **machine learner** uses a training data set that consists of (input, output) examples to induce a set of rules that describes the concepts as contained in the data. That is, the learners learn from the data using a supervised learning approach [5]. The machine learner architecture consists of five components [8].

– The *learning element* contains one of three machine learning techniques that generates sets of propositional rules. The techniques are the BRAINNE approach that extracts rules from a training artificial neural network [6], the CN2 [2] rule induction algorithm that induces rules from training examples, as well as the C4.5Rules algorithm that produces a pruned set of rules that is derived from a decision tree [4].

– The *performance element* controls, monitors and guides the progress of the *learning element*. The *performance element* accepts new examples from the environment and presents the examples to the *learning element* or *critic*. It controls the communication of the rules and quality measures, via the *knowledge-base*, to the environment and the other learners.

– The *critic* contains the evaluation procedure that evaluates the performance of the learner against a set of predetermined quality measures. The CN2 evaluation process is used to evaluate the rules [2]. The *critic* receives the appropriate examples from the *performance element* and evaluates the rules that were formed by the *learning element*.

– The *data generator* is needed to generate new examples that may lead to an improvement in the learner's performance. The new training examples are produced from high quality rules as obtained from other learners.

– The communication element consists of a *knowledge-base* that is used for inter-learner communication. That is, a knowledge-based approach is used to facilitate the transfer of knowledge (rules and quality measures) between learners.

The computational realization of the **human learner** essentially consists of three components, namely a *performance element*, a *knowledge-base* and a *critic*. These three components correspond to those of a machine learner, as described above. The initial knowledge of the human subject is acquired by the knowledge engineer, who interviews the human expert [9]. The expert knowledge is expressed by IF-THEN rules that are placed in a *knowledge-base.* This *knowledge-base* thus reflects the human's current knowledge of the problem at hand. The *critic* contains the CN2 evaluation procedure that is used to evaluate the quality of the human learner's knowledge against one or more sets of examples. The task of the *performance element* is to control and monitor the communication of the rules and quality measures, as contained in the *knowledge-base*, with the other members of the CILT team, as well as to supply the *critic* with new examples.

C. Learning program

The learner program describes the procedures to facilitate cooperative learning. The learners operate in one of three episodes, namely the individual learning, cooperative learning and evaluation episodes [8]. Each learner's learning process is controlled by the *performance element*, which controls and monitors the progress of the *learning element*. Note that this is an iterative process, rather than a sequential one.

– **Individual learning**. During the individual learning episode, the *learning element* of each machine learner creates an initial set of rules that are placed in the *knowledge-base*. Here, the learners do not communicate with one another, but form an initial independent perception of the problem at hand. The *performance element* receives the training examples from the environment and presents it to the *learning element*. Next, the *learning element* proceeds to form sets of rules. The learner forms a number of rule sets by re-iterating the learning process using different internal parameter values.

The knowledge engineer interviews the human subject in order to obtain his perspectives on the concepts to be learned. The knowledge engineer re-phrases the concept descriptions from the results obtained from the interview. These concept descriptions are verified by the human subject and are placed in the human learner's *knowledge-base*, which has the same structure as that produced by the machine learners.

– **Evaluation of initial rule sets.** The evaluation episode is used to determine the quality of a rule set produced by a learner's learning element. Each learner's performance element receives the examples from the environment and presents it to the *critic*, which evaluates the quality thereof. Two measures are of importance. The rule set accuracy refers to the percentage of instances correctly classified by one or more rule contained in the rule set. The rule accuracy denotes the percentage of the cases that was correctly classified by a single rule. For each machine learner, the best rule set, i.e. the one with the highest accuracy against the previously unseen test examples, is selected and placed in the learner's *knowledge-base*.

The human subject improves the quality of his results by modifying the human learner *knowledge-base* until he is satisfied that the *knowledge-base* reflects his current knowledge and is of a high quality.

– **Cooperative learning**. The aim of the cooperative learning episode is to improve the overall quality of the individual results. The input to this episode is the individual *knowledge-bases* that reflect the knowledge as obtained during initial individual learning. Each learner that participates in cooperative learning queries the *knowledge-bases* of the other team members to obtain a NewRule list which contains those high quality rules that it has missed. A high quality rule has an individual rule accuracy that is higher than a predetermined rule accuracy threshold. A machine learner uses each of the rules in its NewRule list to generate a new set of training examples that are used to re-iterate the individual learning and evaluation steps. This process continues until no new rules can be generated or a predetermined period of time has elapsed.

The human subject can improve his initial knowledge by accessing the other team members' rules via the human learner's graphical user interface (GUI). In this way, the human subject may obtain new insights into the problem at hand. The expert's participation is, however, optional. Experience shows that most human experts do not wish to actively participate in the cooperative learning process [7, 9]. This is mainly due to time constraints as well as the human expert's initial skepticism of the machine learning process. Usually, the human experts prefer to obtain the results of the automated process and to verify their beliefs against the machine learning team's *knowledge-bases*.

After cooperative learning has been completed, the knowledge fusion step is executed. In this step, the individual *knowledge-bases* of both the machine and human learners are merged into one. A rule pruning algorithm, which removed redundancies from the rule set, is applied to the team *knowledge-base*. This integrated *knowledge-base* therefore contains the results of team learning.

– **Evaluation of results of cooperative learning**. Finally, the resulting sets of rules that were produced during cooperative learning are evaluated against a new set of previously unseen examples, the so-called validation set. The individual and team results are evaluated by considering the overall rule set and rule accuracies as well as the improvement in these accuracies. A high quality rule set's accuracy should be at least as high as the rule set accuracy threshold, which is equal to the average rule set accuracy as obtained during the initial individual learning process.

This section described the learning approach that is followed when constructing a team *knowledge-base* from a

data repository. The next section discusses the results when applying this method to the NRT Audit data.

IV. Industry Technology Base Survey

The ***Industry Technology Base*** NRT survey was conducted to determine the appropriateness of the technologies used by private firms and public corporations within 17 technology-driven business sectors of South Africa. Data supporting the survey were obtained from 313 significant companies by means of questionnaires and buy-in interviews. The companies were grouped into two sectors, namely **continuous process** and **discrete products** industries, as depicted in Table 1. The **continuous process** sector refers to those industries that produce commodities that are measured in units of volume, mass, length and area, *for example gold, water or chemicals*. These products have well established international benchmark prices and clear product specifications. **Discrete product** industries are those industries that sell their outputs as individual items, *e.g. cars, computers or televisions*. These products are complex to manufacture, produced as an assembly of many different components and frequently made from many different materials.

Each organisation was described in terms of two main aspects. Firstly, the key product lines sold by an organisation were considered with respect to their position in the value chain hierarchy. These positions denoted the unit size of the output, i.e. whether it was a user system, product system, product, sub-products, component or "raw" material. *For example, consider a computer network. Organisations which produce a user system supply the computers, the network, the software as well as the user training capabilities. A product system includes the computer and networking capabilities. An example of a product is a personal computer, which consists of a number of sub-products, i.e. the monitor, keyboard, etc. The sub-products are built up from components (chips, cables, etc.) that have been manufactured from raw material.*

Secondly, an organisation was characterised in terms of the type of technologies that ensures that the organisation's outputs are sustainable in the market (i.e. the so-called key drivers of technology). The technology drivers belonging to the organisations are categorised under four different types, namely product, process, support or informational capabilities. Product technologies are those that are used to produce the outputs. *For example, the gold mining industry uses rock drilling and material separation technologies*. The process technologies are those technologies that are used to produce the final output. *For examples, the gold mining industry uses gold refinement equipment to produce a gold bar. A chocolate manufacturing factory uses wrapping machines to wrap the chocolate bars.* The support and informational technologies are mainly used for after sales support and marketing. These technologies are therefore usually IT-based.

The development stage of each of the technologies used by an organisation was also recorded. This indicates whether the development stage was *emergent*, *pacing*, *key* or *base*. Base technologies are those technologies an organisation requires in order to perform its core business. *For example, an organisation needs a payroll system and an accounting system*. Key technologies refer to the technologies the organisation uses to produce outputs. *For example, a gold mine needs to use gold refinement equipment.* Pacing equipment are well-known and well-established, but may not be used by an organisation due to financial or socio-economic constraints. *For example, a gold mine may, in order to save jobs, decide not to use time-saving rock drilling equipment.* Emergent technologies are new technologies that have not been widely established. Usually, organisations use this technology to gain a competitive advantage. *For example, the gold mine may use cutting edge technologies to speed-up the gold refinement process.*

The application of the KDD learning process to the ***Industry Technology Base*** data is discussed next.

A. Learning process

The cooperative learning team consisted of four learners, namely CN2, C4.5, BRAINNE learner as well as a human expert. Each learner received the full training set and executed the inductive learning step, followed by the evaluation against the training set and the test set. Of the 313 original organisations surveyed, only 249 instances were used, due to incomplete data. The training set consisted of 167 randomly chosen instances, the test set contained 42 instances and the validation set contained the remaining 40 instances. The best rule sets were selected and were placed, together with the quality measures, in the three machine learner's individual *knowledge-bases*. The human expert was interviewed and his knowledge was translated into nine rules. These rules were placed in an individual *knowledge-base* and subsequently evaluated using the CN2 evaluation function.

TABLE 1
CLASSIFICATION OF TWO BUSINESS SECTORS

CONTINUOUS PROCESS	Agriculture, Mining, Base metal Pulp and paper, Power generation Petrochemicals, Glass and non-metallic Food and beverage, Water
DISCRETE PRODUCT	Rubber and plastics, Civil construction Textiles and footware Electrical and electronic Medical and pharmaceutical Automotive and transport, Defense Metal products and machinery

TABLE 2.
ACCURACIES AND RULE SET SIZES OF FOUR LEARNERS AGAINST THE TEST SET

OVERALL RULE SET ACCURACY			
CN2	C4.5	BRAINNE	Expert
81%	83.3%	71.4%	88.1%

NUMBER OF RULES IN KNOWLEDGE-BASE			
CN2	C4.5	BRAINNE	Expert
14	11	5	9

Table 2 shows the initial rule set accuracies and sizes of the four learners, as evaluated against the test set. The rule set produced by the Human Expert had the highest overall accuracy (88.1%), followed by the C4.5 learner (83.3%). The BRAINNE learner failed to find a highly accurate set of rules, with an overall rule set accuracy of only 71.4%. The average rule set accuracy was 80.1% and the average accuracy of the individual rules was 61.2%. These values were used as quality threshold values. The evaluation of the individual rule sets showed that the learners learned accurate rules for the **discrete product** class, but failed to find accurate rules for the **continuous process** class. The aim of cooperative learning was therefore to improve the individual results and to produce sets of informative rules that describe the **continuous process** class.

Next, the three machine learners executed the co-operative learning episode.[1] Each team member queried the *knowledge-bases* of the other team members to find high quality rules that are related to the learner's own low quality rules, as follows:

– The CN2 learner produced five low quality rules during the individual learning episode. Two of the low quality rules were classified as misconceptions and therefore removed from the knowledge base. One low quality rule overlapped with a high quality C4.5 rule. The learner also obtained a high quality rule from the BRAINNE learner. These two rules were placed on the CN2 learner's NewRule list.

– The C4.5 learner produced four low quality rules describing the **discrete product** class. The C4.5 learner obtained two high quality rules, which were placed on the C4.5 NewRule list, from the CN2 team member that overlapped the four low quality rules.

[1] *Note that the human expert did not participate in the cooperative learning process. This was mainly due to scheduling and time constraints. The rule set created by the human expert was, however, used during knowledge fusion in the validation episode.*

TABLE 3.
RULE SET ACCURACIES AND SIZES AFTER KNOWLEDGE FUSION

LEARNER	CN2	C4.5	BRAINNE	ML Team	ML Team & Human Expert
Accuracy	85.0%	72.5%	67.5%	82.5%	87.5%
Number of rules	9	4	3	5	11

– The BRAINNE learner produced two low quality rules during the individual learning episode, one describing each class. The BRAINNE learner obtained two high quality rules, one from the CN2 and C4.5 team members each. These rules were added to the NewRule list.

These NewRule lists were used to re-iterate the cooperative learning process. Table 3 shows the resultant *knowledge-bases* at the completion of cooperative learning.

The *knowledge-base* created by the fusion of the machine learning team's rule set with the human expert's rule set had the highest overall accuracy and contained eleven rules. Here, the diverse learning styles of the four learners lead to the creation of a high quality, informative *knowledge-base*. The **continuous process** rules, as generated by the team during knowledge fusion, had an accuracy of 83.3% against the validation set. For this class, the CN2 learner produced the second highest rule set, with an accuracy of 58.3%. This implies that, by cooperation, the team's description of the **continuous process** sector improved by 25%. The rules describing the **discrete product** class were 100% accurate against the validation set, thus capturing the knowledge describing this sector. Cooperative learning thus improved the generalisation abilities of the team and the quality of the knowledge contained in the team *knowledge-base*.

B. Discussion

This section highlights the main findings as contained in the team *knowledge-base* and discusses how the rules obtained through cooperative learning differ from the human expert's perspectives. The human perspectives were obtained through interviewing the human expert as well as consulting the Synthesis Report, as introduced in Section 2.

The human expert was convinced that the process and product capabilities, when applied in the two sectors, should be identical. The rules obtained from the data did not support this perspective. Rather, it indicated that there is a significant difference in the application of process capability technologies. Recall that this technology type refers to the technologies and competencies directly related to the manufacturing process. The results showed that 97.5% of the **continuous process** industries apply *process* technology

types. On the other hand, only 42% of the **discrete product** industries use technologies of this type. This implies that the **discrete product** industries do not use process technologies when creating their products. Rather, the discrete product industries employ manual labour to produce the final product. This approach leads to job creation, but causes some South African manufactured goods to be too highly priced when compared to other emerging markets. The results did, however, confirm the expert's opinion with respect to the use of product capability technologies, i.e. those technologies used to produce the final product That is, the **discrete product** and the **continuous process** industries both used product capability technologies to produce their outputs.

The knowledge as contained in the team *knowledge-base* showed the textile and footwear industry to be unique when compared to the other **discrete product** industries. The textile and footwear industry did not identify any key technology driver. However, the organisations that participated in the Audit indicated that exports should increase by the year 2000, due to the application of key technologies. The organisations did not indicate how this would be accomplished without the appropriate key technology drivers. It is interesting to note that the South African textile and footwear industry is currently experiencing serious difficulties and that a number of factories had to close down. This industry has difficulty competing internationally, since the production costs when using manual labour are higher than their international competitors.

The human expert found that technologies in the *key* stage of sophistication are more dominant in the **discrete product** industries. On the other hand, technologies in the *base* or *emergent* stage of sophistication are more dominant in the **continuous process** industries. A total of 26% of the **continuous process** industries employed technologies in the *emergent* technological stage of sophistication, compared to only 6% of the **discrete product** industries. In addition, 62% of the **discrete product** industries used technologies in the *key* technological stage of sophistication, compared to only 46% of the **continuous process** industries. This indicates that the rules, as produced during cooperative learning, supported the human expert's findings. The international prices and manufacturing specification of the products produced by the **continuous process** industries are usually fixed. *For example, the gold price is internationally determined and the process of producing gold is well established.* **Continuous process** industries, aiming to compete internationally, should therefore use *emergent* technologies in order to minimise their production costs. Therefore, these industries combine the basic technologies together with "cutting-edge" technologies for a competitive advantage.

This section discussed the application of the KDD approach to a subset of the NRT Audit warehouse. The results obtained indicate that the NRT Audit data warehouse contains a wealth of usable information that can be used to aid the decision makers when creating the Knowledge and Technology Policy Framework for South Africa. It is, however, important to remember that it is not only the knowledge of the experts that will ensure the successful compilation of a policy framework, but also the way in which the policy forming process is conducted. The policy forming process should be socially constituted in order to address the issues already mentioned in Section 2.

V. Conclusion

Knowledge discovery from data is an important topic of research, since it addresses the problems associated with the traditional knowledge acquisition process. This paper showed how the KDD approach was applied to the South African NRT data warehouse. Results indicate that this approach may be used to verify the findings of human experts. Importantly, the use of knowledge discovery from data may lead to obtaining new insights into problem domains.

The NRT Audit data warehouse case study indicated that the use of the KDD process when constructing a *knowledge-base* provides a valuable tool in developing and building the technology policy framework, due to the following reasons:

– Domain experts have seen that they can successfully verify their results against the data. This helped in battling the skepticism that experts traditionally reveal towards the machine learning process. This made it easier to obtain their participation, which is of paramount importance when conducting machine learning.

– The results as contained in the team *knowledge-base* confirmed certain pre-empted ideas from experts, but also showed where the pre-empted ideas were wrongfully made. The approach can therefore be used to ensure that decisions are taken according to correct assumptions.

– The KDD process provides a tool that will enable government to make sense of the large amount of data received from ground level. The socio-economic threats of the incorrect application of a Knowledge and Technology policy in South Africa can, if care is not taken, widen the gap between the economic 'haves' and 'have-nots'. With this in mind, it is very important to consider the way in which policy will be formulated in South Africa. These inputs can sensibly be used to ensure that the policy forming process (or suggested Knowledge and Technology Policy Framework) is not top-down driven, but rather a bottom-up approach.

Further work includes the extension of our approach to include additional learners as well as the continued mining of the data as contained in the NRT data warehouse. In particular, the active participation of the human expert *during*

the actual cooperative learning process should prove worthwhile.

REFERENCES

[1] P Adriaans and D Zantinge, 1997. *Data mining*, Addison-Wesley, Harlow; England.

[2] P Clark and T Niblett, 1989. The CN2 Induction Algorithm, *Machine Learning*, 3, pp.216-283.

[3] T Mitchell, 1997. *Machine Learning*, McGraw-Hill, New York: USA

[4] JR Quinlan, 1994. *C4.5: Programs for Machine Learning*, Morgan Kaufman, California: USA.

[5] SJ Russell and P Norvig, 1995. *Artificial Intelligence: A Modern Approach*, Prentice-Hall, New Jersey: USA.

[6] S Sestito and TS Dillon, 1994. *Automated Knowledge Acquisition,* Prentice-Hall, Sydney: Australia.

[7] E Turban, 1997. *Decision support systems and expert systems* (Fifth Edition), Prentice-Hall, London: UK.

[8] HL Viktor, 1999. *Learning by Cooperation: An Approach to Rule Induction and Knowledge Fusion,* PhD dissertation, Department of Computer Science, University of Stellenbosch, Stellenbosch: South Africa.

[9] HL Viktor, 1999. Combining Humans and Machine into a Cooperative Multi-agent Learning System, *International Journal of Continuous Engineering Education and Life-long Learning,* Inderscience Enterprises, Geneva: Switzerland.