
**The use of a multi-agent learning system to analyse embedded
context in qualitative data for decision-making.**

by

Heidi Arndt

Submitted in fulfilment of the requirements for the degree

MAGISTER COMMERCII (Informatics)

in the faculty of Economic and Management Sciences at the

University of Pretoria

PRETORIA

June 2000

RECEIVED
UNIVERSITY OF PRETORIA
2000-06-15
HEIDI ARNDT
MAGISTER COMMERCII

I declare that

Acknowledgements

The use of a multi-agent learning system to analyse embedded context in qualitative data for decision-making

I would hereby like to express my sincere thanks and gratitude towards:

Is my own work and that all sources that I have used or quoted have been indicated and acknowledged by means of complete references.

• My parents for their encouragement and support.

• My supervisor, Dr. [Name], for his guidance, advice and support.

• The National Research Foundation (NRF) for their financial support of this research project.

• The Department of Arts, Culture, Science and Technology for their support of the National Research Foundation's research project.

Acknowledgements

I would hereby like to express my sincere thanks and gratitude towards:

- Prof HL Viktor for her leadership and assistance.
- My parents for their encouragement and interest.
- A special word of thanks to my husband, Wikus, children, Stefan and Riki, for your love and support.
- The financial assistance of the National Research Foundation (NRF) towards this study is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the National Research Foundation.
- And the Department of Arts, Culture, Science and Technology for making the data and reports of the National Research and Technology Audit available. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the Department of Arts, Culture, Science and Technology.

Abstract

The use of a multi-agent learning system to analyse embedded context in qualitative data for decision-making

Candidate: H Arndt
Study Leader: Prof HL Viktor
Department: Informatics
Degree: M. Com. (Informatics)

A number of studies have shown that the success of knowledge discovery from data, with an intelligent data analysis tool, is dependent on the combination and integration of individual data mining techniques. The aim of this study was to determine whether an intelligent data analysis tool could successfully be used to analyse the context embedded in a real world data repository. Although the data repository contained both quantitative and qualitative measures, the study only focussed on the qualitative aspects of the data. For example, organisations were characterised in terms of the key technologies that made their products sustainable in the market rather than its market share.

The intelligent data analysis tool was based on a multi-agent learning system that consisted of learning agents or so-called learners grouped into learner teams. A learner team included data mining techniques as well as human learners. These learners interacted with one another and the environment. The interactions between the learners involved learning in a co-operative inductive learning team. This was accomplished by team members sharing their knowledge, i.e. the rules they have acquired during the learning process. The knowledge acquired by each individual learner, as well as the team's knowledge were stored in separate knowledge bases.

The intelligent data analysis tool was evaluated against a data repository developed as part of the National Research and Technology (NRT) Audit conducted for the Department of Arts, Culture, Science and Technology of the South African Government. The results of the cooperative learner teams were verified by the active participation of a human expert, as well as against a synthesis report. This report, which was another major output of the NRT Audit, contained findings of experts that described the current state of science and technology in South Africa. Also, it outlined certain trends that were based on the data collected during the NRT Audit.

Experimental results indicated that the intelligent data analysis tool could be applied successfully to a real-world application. It was concluded that the inclusion of a human learner makes a substantial contribution to a multi-agent learning system. The intelligent data analysis tool can be successfully used by human experts to verify their findings and therefore assist them in gaining confidence in their own interpretation of the data. The results obtained from the application of the tool differed from the opinions of the human experts in some instances, indicating pre-conceived ideas that were erroneously made. The human experts indicated that their inclusion in the learning process was a valuable learning experience.

Opsomming

Die gebruik van 'n multi-agent leerstelsel vir die analise van verskuilde konteks in kwalitatiewe data om besluitneming te ondersteun

Kandidaat: H Arndt

Studieleier: Prof HL Viktor

Departement: Informatika

Graad: M. Com. (Informatika)

Verskeie studies het aangetoon dat die sukses van 'n kennisontdekkingstrategie uit 'n data omgewing, met 'n intelligente data analise werktuig, afhanklik is van die kombinasie en integrasie van individuele datamyn tegnieke. Die doel van hierdie studie is om te bepaal of 'n intelligente data analise werktuig toegepas kan word in 'n werklike situasie om die konteks verskuil in 'n databank te kan analiseer en sodoende die besluitnemers te ondersteun. Die databank bevat beide kwalitatiewe sowel as kwantitatiewe maatstawwe. Hierdie studie het gekonsentreer op die kwalitatiewe aspekte. Byvoorbeeld, 'n organisasie is beskou in terme van sleuteltegnologieë wat die firma lewensvatbaar in die mark maak, eerder as die organisasie se markaandeel.

Die intelligent data analise werktuig is gebaseer op 'n multi-agent leerstelsel wat bestaan uit leeragente wat in leerspanne gegroepeer is. 'n Leerspan bestaan uit beide datamyn tegnieke, sowel as kundige individue, wat met mekaar en die omgewing saamwerk. Die interaksie tussen die leeragente behels koöperatiewe leer wat plaasgevind het in spanne en bekend staan as koöperatiewe induktiewe leerspanne. Dit was bewerkstellig deurdat spanlede hulle kennis, die stel reëls wat hulle gegenereer het gedurende die leerproses, met mekaar uitruil. Die kennis wat elke individuele leerder, sowel as die span as geheel ontdek het, is in aparte kennisbasisse gestoor.

Die intelligente data analise werktuig is gebruik om die databank, ontwikkel as deel van die Nasionale Navorsing en Tegnologie (NRT) Oudit van die Departement Kuns, Kultuur, Wetenskap en Tegnologie van die Suid-Afrikaanse Regering, verder te ontleed. Die resultate van die koöperatiewe induktiewe leerspanne is getoets deur die aktiewe deelname van 'n kenner in die gebied, sowel as teen die sintese verslag wat nog 'n uitset van die NRT Oudit was. Hierdie verslag bevat die bevindinge van gebiedskenners wat die huidige stand van die wetenskap en tegnologie in SA beskryf, tesame met sekere tendense bepaal vanaf die data wat gedurende die Audit versamel is.

Die resultate toon dat die intelligente data analise werktuig wel suksesvol in 'n werklike situasie toegepas kan word. In hierdie toepassing het die menslike leerder 'n beduidende bydrae gemaak tot die multi-agent leerstelsel. Die strategie kan gebruik word om die gevolgtrekkings van die kenners betrokke in die analise te bevestig en hulle sodoende te help om vertroue in hulle eie interpretasie van die data op te bou. Die strategie het ook soms verskil van die menings van die betrokke kenners en het daardeur foutiewe aannames uitgewys, wat weer vir die individue 'n waardevolle leerervaring was.

TABLE OF CONTENTS

CHAPTER 1 1

INTRODUCTION 1

1.1 Problem statement..... 3

1.2 Research approach..... 6

1.3 Thesis outline..... 8

CHAPTER 2 9

CO-OPERATIVE INDUCTIVE LEARNING IN A MULTI-AGENT LEARNING SYSTEM. 9

2.1 Co-operative inductive learner teams..... 10

2.2 Multi-agent learning system 12

2.3 Modelling a co-operative inductive learner team as a multi-agent learning system..... 14

2.4 Learning in the CILT-MAL system 18

 2.4.1 Individual learning phase 20

 2.4.2 Co-operative learning episode..... 21

 2.4.3 Evaluation and knowledge fusion episodes..... 23

2.5 Langley’s framework for machine learning..... 24

 2.5.1 Key aspects of the environment and performance task measures 26

 2.5.2 Key aspects of the knowledge base 27

 2.5.3 Key aspects of a learning mechanism..... 28

 2.5.4 CN2 – A rule induction learning agent..... 30

 2.5.4.1 Environment and performance task measures..... 30

 2.5.4.2 The knowledge base and performance element 31

 2.5.4.3 The learning mechanism..... 32

2.5.5	C4.5 – A decision tree learning agent.....	33
2.5.5.1	Environment and performance task measures.....	33
2.5.5.2	The knowledge base and performance element	33
2.5.5.3	The learning mechanism.....	36
2.5.6	BRAINNE– an artificial neural network learning agent with rule extraction algorithm..	38
2.5.6.1	Environment and performance task measures.....	38
2.5.6.2	The knowledge base and performance element	39
2.5.6.3	The learning mechanism.....	40
2.5.7	Human learner– learning agent	40
2.5.7.1	Environment and performance task measures.....	40
2.5.7.2	The knowledge base and performance element	41
2.5.7.3	The learning mechanism.....	43
2.6	Conclusion	44
CHAPTER 3	47
NATIONAL RESEARCH AND TECHNOLOGY AUDIT	47
3.1	Survey overviews.....	49
3.1.1	Survey of the scientific and technological infrastructure.....	51
3.1.2	Survey of human resources and skills in science, engineering and technology.....	51
3.1.3	Survey of scholarship, research and development.....	51
3.1.4	Survey of the technology base of the South African business sector	52
3.1.5	Survey of research and training equipment in South Africa.....	53
3.2	Summation of the findings.....	53
3.2.1	Quality of human resources.....	54
3.2.2	Relevance of the science and technology system.....	55
3.2.3	Effectiveness and efficiency of the science and technology system	55
3.3	The two learning tasks	56
3.4	Conclusion	57
CHAPTER 4	58

CLASSIFICATION TASK: GROUPING ORGANISATIONS INTO CONTINUOUS PROCESS AND DISCRETE PRODUCT BUSINESS-CLUSTERS.	58
4.1 Task description	59
4.2 Data pre-processing.....	59
4.3 Initial exploratory results	61
4.4 Experimental method and evaluation criteria	63
4.5 Learning of the classification task by the machine learner team	64
4.5.1 Individual learning phase	64
4.5.2 Co-operative learning episode.....	65
4.5.2.1 The CN2 co-operative learning episode	66
4.5.2.2 The C4.5 co-operative learning episode	69
4.5.2.3 The BRAINNE co-operative learning episode.....	72
4.6 Learning of the classification task by the human learner team.....	76
4.6.1 Individual learning phase	76
4.7 Validation evaluation episode and knowledge fusion	78
4.8 Discussion	80
4.9 Conclusion	82
CHAPTER 5	84
PROBLEM SOLVING TYPE TASK: DISCOVERING TRENDS IN THE HUMAN RESOURCES FOR RESEARCH AND TECHNOLOGY.....	84
5.1 Task description	85
5.2 Data pre-processing.....	85
5.3 Initial exploratory results	88
5.4 Experimental method and evaluation criteria	89

5.5 Learning of the problem solving type task by the combined machine-human learner team	91
5.5.1 Individual learning phase	91
5.5.2 Co-operative learning episode	94
5.5.2.1 The CN2 co-operative learning episode	94
5.5.2.2 The C4.5 co-operative learning episode	97
5.5.2.3 The human learner co-operative learning episode	97
5.6 Validation episode and knowledge fusion	98
5.7 Discussion	100
5.8 Conclusion	102
CHAPTER 6	104
SUMMARY AND CONCLUSION	104
6.1 Summary	105
6.2 Concluding remarks.....	106
BIBLIOGRAPHY	109

INDEX OF FIGURES

FIGURE 1: RESEARCH APPROACH..... 7

FIGURE 2: CO-OPERATIVE LEARNER TEAM MODELLED AS A MULTI-AGENT
LEARNING SYSTEM..... 15

FIGURE 3: MODEL OF A CO-OPERATIVE INDUCTIVE LEARNING AGENT USING THE
INDUCTIVE MACHINE LEARNING ARCHITECTURE [VIKTOR 1999] 16

FIGURE 4: THE LEARNING PROCESS OF THE CILT-MAL SYSTEM..... 18

FIGURE 5: LANGLEY’S MACHINE LEARNING FRAMEWORK 25

FIGURE 6: BRAINNE CONCEPT DESCRIPTION LANGUAGE..... 40

FIGURE 7: BUSINESS SURVEY ENTITY RELATIONSHIP DIAGRAM..... 60

FIGURE 8: HUMAN RESOURCES DATA STRUCTURE..... 85

INDEX OF TABLES

TABLE 1: VEHICLE DESCRIPTIONS	11
TABLE 2: KEY ASPECTS OF THE LEARNING AGENTS MODELLED USING LANGLEY'S MACHINE LEARNING FRAMEWORK	46
TABLE 3: INDUSTRY SECTORS.....	53
TABLE 4: SAMPLE DATA FROM THE BUSINESS SECTOR PROFILE DATA SET	62
TABLE 5: ACCURACIES AND RULE SETS AFTER THE INDIVIDUAL MACHINE LEARNER EPISODES	65
TABLE 6: ACCURACIES AND RULE SETS AFTER THE CO-OPERATIVE LEARNING EPISODE.....	75
TABLE 7: ACCURACIES AND RULE SETS AFTER THE INDIVIDUAL LEARNING PHASE OF BOTH TEAMS	78
TABLE 8: ACCURACIES AND RULE SETS AFTER VALIDATION AND KNOWLEDGE FUSION.....	79
TABLE 9: FINAL ACCURACIES AND RULE SETS OF FUSED KNOWLEDGE BASES.....	79
TABLE 10: SCARCE DISCIPLINES OF HUMAN RESOURCES IN RESEARCH AND TECHNOLOGY	87
TABLE 11: SAMPLE DATA FROM THE EMPLOYEE-PROFILE DATA SET	88
TABLE 12: STRENGTHS AND INTERESTINGNESS OF TRENDS AFTER THE INDIVIDUAL LEARNING PHASE.....	92
TABLE 13: LOW QUALITY TRENDS GENERATED BY THE CN2 LEARNER	95
TABLE 14: TRENDS GENERATED BY THE CN2 LEARNER DURING THE CO-OPERATIVE LEARNING EPISODE	96
TABLE 15: RESULTS OF THE CO-OPERATIVE LEARNING EPISODES.....	98