



Ain Shams University  
Faculty of Engineering  
Computer and Systems Engineering Department

# **Knowledge Discovery in Intelligent Tutoring Systems**

A Thesis

Submitted in Partial Fulfillment of the Requirements of the  
Degree of Master of Science in Electrical Engineering  
(Computer and Systems Engineering Department)

Submitted By

**Sally Sameh Abd El Ghaffar Attia**

B. Sc., Electrical Engineering  
(Computer and Systems Engineering)  
Ain Shams University, 1998

Supervised by

**Prof. Dr. Hani M. K. Mahdi**

**Dr. Hoda K. Mohamed**

Cairo, Egypt

November, 2004



Faculty of Engineering

---

## **Examiners Committee**

**Name** : Sally Sameh Abd El Ghaffar Attia

**Thesis** : Knowledge Discovery in Intelligent Tutoring Systems

**Degree** : Master of Science in Electrical Engineering

<b>Name, Title and Affiliation</b>	<b>Signature</b>
<b>1. Prof. Dr. Mohamed Zaki Abdel-Megeed</b> Head of Systems Engineering and Computer Dept. Faculty of Engineering, Al-Azhar University, Cairo	.....
<b>2. Prof. Dr. Abdel-Moneim A. Wahdan</b> Head of Computer and Systems Engineering Dept. Faculty of Engineering, Ain Shams University, Cairo	.....
<b>3. Dr. Hoda Korashi Mohamed</b> Computer and Systems Engineering Dept. Faculty of Engineering, Ain Shams University, Cairo (Supervisor)	.....

**Date:21/11/2004**

## **Acknowledgements**

I owe a special acknowledgment to Prof. Hani M. K. Mahdi for spending a lot of his precious time reading my work, writing his comments, suggesting improvements and discussing technical points with me specially during preparing my first paper with him and Dr. Hoda to be published in ICEEC'04 in Cairo.

I am deeply grateful to Dr. Hoda K. Mohammad for supporting me during the process of writing the thesis, reading it several times and giving me her valuable comments and advice and helping me to finish the thesis in time.

I will never be able to thank my parents enough for supporting me during all my life, especially during the last period of preparing the thesis. They gave me hope and help when I really needed them and had to bear a lot so that I could finish this work. I would like to dedicate this work to them and also to my sister who helped me in drawing the figures.

I am very grateful to my husband for helping me processing the thesis whenever I needed his help and for bearing that I was always busy during the last period. I would also like to dedicate this work to my lovely two kids, Youssef and Nour.

# **Abstract**

**Sally Sameh Abd El Ghaffar Attia**

**Knowledge Discovery in Intelligent Tutoring Systems**

**Master of Science dissertation**

**Ain Shams University, 2004**

---

Knowledge discovery through data mining aims at searching for meaningful information like patterns and rules in large volumes of data. Our objective is to discover useful knowledge in Intelligent Tutoring Systems (ITS). These are tutoring systems which offer the ability to respond to individualized student needs. A major challenge for today's schools is to create individualized instruction in inexpensive, expandable ways. To achieve this goal, educational systems must adapt to each student by mining student data to determine what a student knows and does not know.

An experiment was conducted over a tutorial for binary relations. Students' answers to questions at the end of the lesson were collected. Data mining using Rough Sets technique was implemented to extract important rules from the data (students' answers) and hence the student can be directed to which parts of the lesson he should take again, thus helping to adopt the tutoring systems to each student individual needs and hence the tutoring systems are called intelligent. Using this knowledge, teaching systems can guide students in the learning process.

In this research, three different approaches of Rough Sets are applied to detect the decision rules. These approaches provide a powerful foundation to discover important structures in data. These approaches are unique in the sense that they only use the information given by the data and do not rely on other model assumptions. We found that, rough sets can be used to understand

large sets of student data, pinpointing problem areas for student learning.

The results obtained were in the form of decision rules that showed what concepts the student understood and which he did not understand depending on which questions he answered correct and which questions he answered wrong. Also some questions of the quizzes were found to be useless. It was concluded that data mining was able to extract some important patterns and rules from the students' answers which were hidden before and which are helpful to both the students and the experts.

In conclusion, this research has provided an in-depth analysis of the use of rough sets approaches as a mining tool in computer-based tutorials. The application of rough sets methods in diagnosing student misconceptions makes a contribution to the fields of computer-based education, fault-tolerant teaching, and data mining.

**Keywords:**

Knowledge Discovery; Data Mining; Intelligent Tutoring Systems; Rough Sets; Fault Tolerant Teaching.

## **Statement**

This dissertation is submitted to Ain Shams University for the degree of Master of Science in Computer and Systems Engineering Department.

The work included in this thesis was done by the author at Computer and Systems Engineering Department, Ain Shams University.

No part of this thesis has been submitted for a degree or qualification at other university or institution.

Date : 21/ 11/ 2004

Signature :

Name : Sally Sameh Abd El Ghaffar Attia

## TABLE OF CONTENTS

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
<b>2</b>	<b>An Overview of Data Mining and Intelligent Tutoring Systems</b>	<b>6</b>
<b>2.1</b>	<b>Knowledge Discovery and Data Mining.....</b>	<b>6</b>
2.1.1	Introduction.....	6
2.1.2	Examples .....	7
2.1.3	Relevance of Data Mining.....	9
2.1.4	Steps to Knowledge Discovery .....	10
2.1.5	Challenges .....	14
2.1.6	Data Mining Outcomes .....	15
2.1.7	Approaches to Data Mining .....	18
2.1.8	Data Mining Techniques and Algorithms .....	19
2.1.9	Decision Support .....	25
2.1.10	Comparison of Data Mining Methodologies .....	25
2.1.11	Accuracy of the Mined Model.....	27
2.1.12	Recent Research in Data Mining.....	28
<b>2.2</b>	<b>Intelligent Tutoring Systems.....</b>	<b>30</b>
2.2.1	Components of Intelligent Tutoring Systems.....	32
2.2.2	Types of Intelligent Tutoring Systems .....	36
2.2.3	The Student Model.....	39
2.2.4	Pedagogical Module .....	43
2.2.5	The Great Debates of ITS and Future Work .....	46
2.2.6	Design of Intelligent Tutoring Systems.....	49
2.2.7	Examples of Intelligent Tutoring Systems.....	50
2.2.8	Conclusion.....	55
<b>2.3</b>	<b>Connection between Data Mining and Intelligent Tutoring Systems.....</b>	<b>55</b>
2.3.1	Data Mining, Teaching, and Assessment .....	57
2.3.2	Adaptive Tutorial Systems and Student Assessment .....	59
2.3.3	Research Objectives.....	61
2.3.4	Adaptive Tutorial Design .....	62
2.3.5	Background of Procedural Knowledge Model .....	65

<b>3</b>	<b>Rough Sets Methodology.....</b>	<b>72</b>
<b>3.1</b>	<b>Rough Sets.....</b>	<b>72</b>
3.1.1	Information System .....	73
3.1.2	Indiscernibility Relation.....	73
3.1.3	Decision Tables .....	74
3.1.4	Approximation of Sets .....	75
3.1.5	Evaluation of Approximation Sets .....	76
3.1.6	Core and Reducts of Attributes .....	78
3.1.7	Discernibility Matrix .....	80
3.1.8	Algorithm for Reduced Relation Generation.....	81
3.1.9	Attribute Core of Decision Table from the Information Point of View ...	86
3.1.10	Decision Rules .....	91
<b>3.2</b>	<b>Modified Rough Sets.....</b>	<b>96</b>
3.2.1	Approximate Decision .....	97
3.2.2	Decision Rules According to Traditional Rough Sets Approach.....	101
3.2.3	Approximate Rules Using Modified Rough Sets .....	102
<b>3.3</b>	<b>Evaluation of the Extracted Decision Rules.....</b>	<b>105</b>
3.3.1	Ambiguity Measure .....	108
3.3.2	Complexity Measure.....	109
<b>3.4</b>	<b>Decision Analysis.....</b>	<b>115</b>
<b>3.5</b>	<b>Rough Sets and Data Mining.....</b>	<b>116</b>
<b>3.6</b>	<b>Comparison with Some Results in Machine Learning.....</b>	<b>117</b>
3.6.1	Rough Sets and Fuzzy Sets.....	117
3.6.2	Hybrid Methods and Systems .....	118
<b>3.7</b>	<b>Applications.....</b>	<b>118</b>
<b>3.8</b>	<b>Rough Sets and Knowledge Discovery.....</b>	<b>119</b>
<b>4</b>	<b>Analysis of Data Samples.....</b>	<b>121</b>
<b>4.1</b>	<b>Background of the NovaNET Network.....</b>	<b>121</b>
<b>4.2</b>	<b>Analysis of Students' Responses for the Binary Relations Tutorial.....</b>	<b>121</b>



<b>4.3</b>	<b>Analysis of Expert Q-matrices.....</b>	<b>130</b>
4.3.1	Expert Q-matrix for the Binary Relations Tutorial Section 1.....	130
4.3.2	Expert Q-matrix for the Binary Relations Tutorial Section 2.....	131
4.3.3	Expert Q-matrix for the Binary Relations Tutorial Section 3.....	133
<b>4.4</b>	<b>Summary of Knowledge Assessment.....</b>	<b>134</b>
<b>4.5</b>	<b>Survey Results.....</b>	<b>135</b>
4.5.1	First Part of the Survey .....	137
4.5.2	Survey Free Response Questions .....	139
4.5.3	Summary of Survey Results .....	142
<b>4.6</b>	<b>Code Implementation.....</b>	<b>143</b>
<b>5</b>	<b>Results and Discussion.....</b>	<b>154</b>
<b>5.1</b>	<b>Fall 2002 Data Set.....</b>	<b>154</b>
5.1.1	Threshold = 0.....	155
5.1.2	Threshold = 1.....	158
<b>5.2</b>	<b>Fall 2001 Data Set.....</b>	<b>162</b>
5.2.1	Threshold = 0.....	162
5.2.2	Threshold = 1.....	165
<b>5.3</b>	<b>General Discussion.....</b>	<b>169</b>
<b>6</b>	<b>Conclusions and Future Work.....</b>	<b>171</b>
<b>Appendix A</b>	<b>Rough Sets Case Studies.....</b>	<b>176</b>
<b>Appendix B</b>	<b>Data Structures in the Implementation.....</b>	<b>185</b>
<b>References</b>	<b>.....</b>	<b>188</b>

## LIST OF TABLES

Table 2-1 Binary Q-matrix Example .....	70
Table 3-1 Inconsistent Decision Table .....	87
Table 3-2 CARS Decision Table .....	95
Table 3-3 Information System .....	99
Table 3-4 Vertical Reduction of the Information System.....	100
Table 3-5 Modified Information System .....	103
Table 3-6 Vertical Reduction of the Modified Information System .....	104
Table 3-7 Values of $ X_i \cap Y_j $ .....	111
Table 3-8(a) Values of RES(i, j) .....	111
Table 3-8(b) Values of COM(i, j) .....	112
Table 3-9 Values of $ X_i \cap Y_j $ .....	113
Table 3-10(a) Values of RES(i, j) .....	113
Table 3-10(b) Values of COM(i, j) .....	114
Table 4-1 Binary Relations Tutorial Section 1 Comparison with Student Exam Grades.....	123
Table 4-2 Binary Relations Tutorial Section 2 Comparison with Student Exam Grades.....	124
Table 4-3 Binary Relations Tutorial Section 3 Comparison with Student Exam Grades.....	125
Table 4-4 Correlation of Binary Relations Tutorial Section 1 Scores with Final Exam Scores .....	128

Table 4-5	Correlation of Binary Relations Tutorial Section 2 Scores with Final Exam Scores .....	128
Table 4-6	Correlation of Binary Relations Tutorial Section 3 with Final Exam Scores .....	129
Table 4-7	Topics in Binary Relations Tutorial, Section 1 .....	130
Table 4-8	Binary Relations Section 1 Expert Q-matrix .....	131
Table 4-9	Topics of Binary relations Section 2 .....	132
Table 4-10	Binary Relations Section 2 Sellers' Expert Q-matrix .....	132
Table 4-11	Topics in Binary Relations Tutorial Section 3 .....	133
Table 4-12	Binary Relations Section 3 Expert Q-matrix .....	134
Table 4-13	Tutorial Survey Questions .....	136
Table 4-14	Tutorial Survey Results for Fall 2002 Data for the Binary Tutorial, by Percentage of the Respondents .....	137
Table 4-15	Answer Vectors (in Octal Reverse) and Number of Students .....	144

## LIST OF FIGURES

Figure 2-1 Some data mining application areas [8] .....	9
Figure 2-2 Knowledge discovery scheme [1] .....	11
Figure 2-3 Steps of knowledge discovery process [1] .....	12
Figure 2-4 How to obtain data mining outcomes [8].....	16
Figure 2-5 Interaction of components in ITS [12] .....	32
Figure 2-6 Intelligent tutoring system chart [11] .....	35
Figure 2-7 Representation of student model [12] .....	41
Figure 2-8 Adaptive tutorial system flow [2] .....	64
Figure 2-9 Student's knowledge versus expert's knowledge [2] .....	65
Figure 4-2 Software design flowchart.....	148
Figure 4-3 Snapshot of the interface .....	149
Figure 4-4 Condition and decision attributes.....	150
Figure 5-1 Core and best reduct when threshold = 0.....	154
Figure 5-2(a) Rules extracted by (RS-method A) when threshold = 0 .....	155
Figure 5-2(b) Rules extracted by (RS-method A) after generalization step.....	156
Figure 5-3 Rules extracted by (RS-method B) when threshold = 0.....	157
Figure 5-4 Rules extracted by MRS when threshold = 0.....	157
Figure 5-5 Core and best reduct when threshold =1 .....	158

Figure 5-6(a) Rules extracted by (RS-method A) when threshold = 1 .....	159
Figure 5-6(b) Rules extracted from (RS-method A) after generalization step.....	160
Figure 5-7 Rules extracted by (RS-method B) when threshold = 1.....	160
Figure 5-8 Rules extracted by MRS when threshold = 1 .....	161
Figure 5-9 Core attributes using information view .....	162
Figure 5-10 Core and best reduct when threshold = 0 .....	162
Figure 5-11(a) Rules extracted by (RS-method A) when threshold = 0 .....	163
Figure 5-11(b) Rules extracted by (RS-method A) after generalization step.....	163
Figure 5-12 Rules extracted by RS-method B when threshold = 0 .....	164
Figure 5-13 Rules extracted by MRS when threshold = 0.....	165
Figure 5-14 Core and best reduct when threshold =1.....	165
Figure 5-15(a) Rules extracted by (RS-method A) when threshold = 1 .....	166
Figure 5-15(b) Rules extracted from (RS-method A) after generalization step.....	167
Figure 5-16 Rules extracted by (RS-method B) when threshold = 1.....	167
Figure 5-17 Rules extracted by MRS when threshold = 1.....	168
Figure 5-18 Core attributes according to information view.....	168

# 1 Introduction

As the world becomes more interconnected, information can be shared between people as never before. While this opens up exciting possibilities for communication and research, it also results in an overwhelming amount of information that each person must process every day in order to function. With each new person joining the global village, the need to develop automated methods to filter, sort, search, and share information in meaningful ways grows exponentially. Therefore, there is a big need for data mining in society today.

Data mining is a set of methods used as a step in the Knowledge Discovery (KD) process to distinguish previously unknown relationships, rules and patterns within large volumes of data. One of data mining tasks is description i.e. to describe databases in terms of patterns which human can understand and make use of [1].

Our schools are facing an even greater need for the benefits that data mining promises to deliver. Both teachers and students are faced with teaching and learning an ever-growing number of ideas all in the same school day as before. Ideally, as more issues are addressed, schools would add more teachers and classes to cover new topics. Instead, schools face teacher shortages, inadequate teacher training, and growing classroom sizes.

At the same time, the typical trend in universities is changing from young on-campus students to diverse professionals of all ages taking classes in the evenings, online, or even at a distance. Colleges and universities are now challenged to deliver quality service to students of widely ranging abilities and backgrounds across all distances. To rise to this challenge, information technology assisted education must pair individuals with expert instruction that is interactive,

adaptive, and accessible. The promise of information technology assisted education is great, but research into methods for the delivery of education with information technology is still a rich, open field [2].

In our research, we are trying to mine databases resulting from Intelligent Tutoring Systems (ITS). These are Computer-based tutoring systems which achieve their intelligence by representing pedagogical decisions about how to teach as well as information about the learner. This allows for greater versatility by altering the system's interactions with students. Intelligent tutoring systems have been shown to be highly effective at increasing student's motivation and performance [3].

We are trying to investigate if the relatively new area of research, termed data mining and knowledge discovery, can be applied to educational problems to achieve our goal of individualized instruction, at hopefully much lower costs than traditional adaptive teaching systems that use knowledge-based models to understand and direct student learning.

Many excellent intelligent tutoring systems exist today. However, the majority of these intelligent tutoring systems require the construction of complex models that are applicable only to a specific tutorial in a specific field, necessitating a large number of experts to create and then test these models on students. Data mining and knowledge discovery, on the other hand, might be applied to the problem of understanding student knowledge, and using this understanding to direct knowledge remediation.

The goal of data mining in ITS is to automatically assess student knowledge of the concepts underlying a tutorial topic, and use this