

## Intelligent Technologies for Medical eLearning

ABDEL-BADEEH M. SALEM

Faculty of Computer & Information Sciences  
Ain Shams University, Abbassia, Cairo, Egypt

### Abstract

In the last few years, new forms of artificial intelligence-based education software have been created that allow the computer to act as an intelligent tutor. Such intelligent tutoring systems (ITSs) can adjust its tutorial to the student's knowledge, experience, strengths, and weaknesses. It may even be able to carry on a natural language dialogue. In addition, automatic generation of exercises and tests is an important feature of such systems. Recently, software vendors and researchers start research to put the ITSs on the Internet for the purpose of distance education through the web. Not only do students receive training at their own sites, but instructors monitor student's progress from a distance, and course authors maintain and update training material across the Internet.

On the other hand, ITSs are complex to build, maintain and face the knowledge-acquisition difficulty. Productivity of ITS development is determined by the efficiency of their knowledge representation techniques and reasoning methodologies. This talk presents some of the intelligent technologies used in developing intelligent medical tutoring/learning systems at Ain Shams University, Cairo, are discussed as well.

### 1. Introduction

The field of artificial intelligence (AI) in education has become the most challenging area in the last several years. It includes the disciplines; cognitive and social psychology, computer science, empirical psychology, software and knowledge engineering. The goal of the field is to deliver computer-based systems (or knowledge-based software) which can be used in real teaching, learning and training situations. Using AI concepts and techniques new forms of educational software can be created that allow the computer to act as an intelligent tutor. Intelligent Tutoring Systems (ITSs) are smart computer tutors and typically have an expert model, student model, instructional module, and interface [1]. ITS must organize their knowledge in a lesson-oriented manner. This organi-

zation must be dynamically adjusted by the system according to student models. Additionally, ITS enhances instructor productivity, enabling them to cope with generating more complex training systems required to provide higher skill levels to today's high-tech trainees. It also provides tailored instruction and remediation, while allowing flexibility in teaching methods, achieving many of the same benefits as one-on-one instruction.

On the other hand, Medical Intelligent Systems (MISs) are concerned with the construction of intelligent software that performs diagnosis and make therapy recommendations. Unlike other medical applications based on other programming methods such as purely statistical methods, MISs are based on symbolic models of disease and their relationship to patient factors. Many types of MISs are in existence today and are applied to different medical tasks, e.g. generation alerts and reminders, diagnosis assistant, therapy critiquing and education [2].

In this paper we focus our discussion around the AI-based technologies which are used in developing medical tutoring/learning systems at Ain Shams University. Section 2 gives a brief overview of the field of AI in education. Section 3 presents a brief overview about the knowledge representation techniques which are usually used in developing knowledge-based systems. Section 4 the main concepts of data mining technology. Section 5 presents machine learning techniques, namely, rough sets and genetic algorithms which are currently used in developing data mining medical systems for thrombosis diseases. The final section contains conclusion.

### 2. AI in education

AI is science and technology and is based on many disciplines such as: computer science, psychology, mathematics, biology, linguistics and engineering. The goal of AI is to develop intelligent software models the human behaviour, i.e. think, see, hear, walk, talk, and also feel. The field covers the areas; action and perception (vision, robotics,

auditory scene analysis), automated reasoning, case based reasoning, cognitive modeling, connectionist models, constraint satisfaction, distributed AI, genetic algorithms, knowledge base technology, knowledge representation, learning, natural language, non monotonic reasoning, planning, qualitative reasoning and diagnosis, reasoning under uncertainty and temporal reasoning.

Based on the analysis of the topics of the World Conferences on "Artificial Intelligence in Education (AI-ED)", which held during the period 1993 - 2001, Table 1 shows the main areas of the AI in education [3].

**Table 1.** Areas of AI in education

- Intelligent tutoring systems.
- Learning environment and microworlds.
- Visual and graphical interfaces.
- Human factor and interface design.
- Intelligent multimedia systems.
- Authoring system and tutoring shells.
- Collaboration tools.
- Training job skills.
- Principles/tools for instructional design.
- Natural language interfaces.
- Knowledge representation.
- Knowledge and skill acquisition.
- Metacognition.
- Social and cultural aspect of learning.
- Cognitive development and errors.
- Student modeling.
- Teaching higher-order thinking skill.
- Theories of teaching.
- Motivation.
- Reading and writing.
- Computer-assisted language learning.
- Evaluation of computer systems
- Assessment of learning outcomes.
- Conceptual change.

### 3. Knowledge representation for ITSs

Knowledge representation techniques offer potentially powerful tools for the development of intelligent learning technologies. The variety of such techniques enabling the design of a robust intelligent tutoring and learning systems. The key to the success of such systems is the selection of the knowledge representation scheme that best fits the domain knowledge and the problem to be solved. That choice is depends on the experience of the knowledge engineer.

A variety of knowledge representation schemes are used including; lists, trees, semantic networks, frames, scripts and production rules. Lists are used to represent hierarchical knowledge. Hierarchical knowledge can also be represented

visually with graphs called trees. Semantic networks use circles called nodes that represent objects or events. The nodes are interconnected with lines called arcs that show relationships. Frames and scripts are two types of schemes dealing with stereotyped knowledge. Frames are used represent facts about objects and events. And details are given in sub-elements called slots. Scripts describe knowledge that is a sequence of events or procedures. Frames and scripts permit a system to infer details of specific common objects and events. Production rules are the most commonly used knowledge representation methods. Production rules are two part statements with a premise and a conclusion. They also may state a situation and corresponding action. Rules take the form of an if-then statement.

### 4. Data mining technology

Data Mining (DM) deals with the discovery of hidden knowledge, unexpected patterns and new rules from large databases [14]. Data mining and knowledge discovery is not a coherent field, it is a dwells upon already well established technologies including data cleaning, data preprocessing, machine learning, pattern recognition, statistics, neural networks, fuzzy sets, rough sets, clustering, etc. Fig. 1 shows the main functional phases of the knowledge discovery process. The preprocessing phase is often referred to as data cleaning. The cleaned data are stored in the warehouse. This is followed by data mining phase and its results are provided to an output generator (visualization) producing reports, action lists, or monitor reports. Each phase is supported by different methodologies. Data mining itself exhibits a plethora of algorithmic tools such as statistics, regression models, neural networks, fuzzy sets and evolutionary model.

The knowledge discovery process is arranged into a stream of steps:

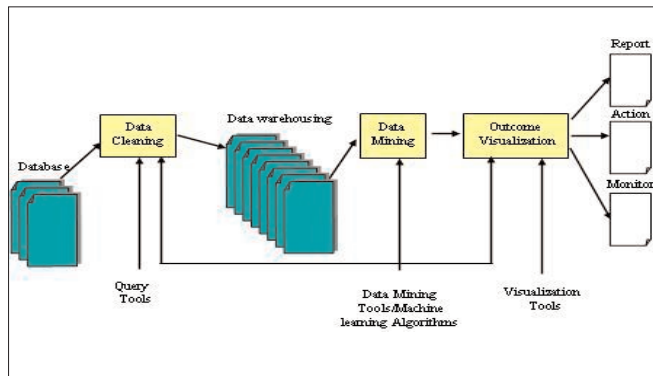
- understanding the domain in which the discovery will be carried out.
- forming the data set, its cleaning, and warehousing.
- extracting patterns, this is essence of DM.
- post-processing of the discovery knowledge
- putting the results of knowledge discovery into use.

Fundamental issues in knowledge discovery arise from the very nature of databases and the objects (data) they deal with. They are characterized as follows:

- huge amounts of data.
- dynamic nature of data.
- incomplete or imprecise data.
- noisy data.
- missing attribute values.

- redundant or insignificant data

**Figure 1.** Phases of knowledge discovery process.



The knowledge discovery process is dynamic, highly interactive, iterative, and fully visualizable. Its main goals are to:

- extract useful reports
- spot interesting events and trends
- support decision-making processes
- exploit the data to achieve scientific, business, or operational goals.

In spite of the diversity of the application areas, there are several common characteristic features:

- availability of massive sets of data
  - high underutilization of data.
  - access domain experts fully familiar with the area becomes crucial during the development of the knowledge discovery system.
  - lack of expertise of the end users (thus knowledge discovery is a greatly welcomed activity).
- Data mining is supported by a host that captures the character of data in several different ways.
- **Clustering:** The key objective is to find natural groupings (clusters) in highly dimensional data. Clustering is an example of unsupervised learning, and it is a part of pattern recognition.
  - **Regression Models:** These originate from standard regression analysis and its applied part known as system identification. The underlying idea is to construct a linear or nonlinear function
  - **Classification:** This concerns learning that classifies data into the predetermined categories. The term originates from pattern recognition, in which a vast number of classifiers have been developed.
  - **Summarization:** This is an approach towards characterizing data via small number of features/attributes. In the simplest scenario one can think of a mean and standard deviations as two extremely compact descriptors of the data. This technique is often applied in an interactive exploratory data analysis and automated report generation.
  - **Link analysis:** It is concerned with determination of relationships (dependencies) between fields in a database. In a particular case we may be

interested in the determination of the correlation between the variables.

- **Sequence Analysis:** This type of analysis is geared toward problems of modeling sequential data. Pertinent models embrace time series analysis, time series models, and temporal neural networks.

## 5. Machine learning-based medical systems

Machine Learning (ML) aims at providing computational methods for accumulating, changing and updating knowledge in intelligent systems, and in particular learning mechanisms that will help us to induce knowledge from examples or data. ML provides methods, techniques, and tools that can help solving diagnostic and prognostic problems in a variety of medical domains. ML is being used for the analysis of the importance of clinical parameters and of their combinations for prognosis, e.g. prediction of disease progression, for the extraction of medical knowledge of outcomes research, for therapy planning and support and for overall patient management [5].

In what follows, we briefly present some of the ML techniques which are currently used in developing medical data mining systems at Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo.

### 5.1 Rough sets

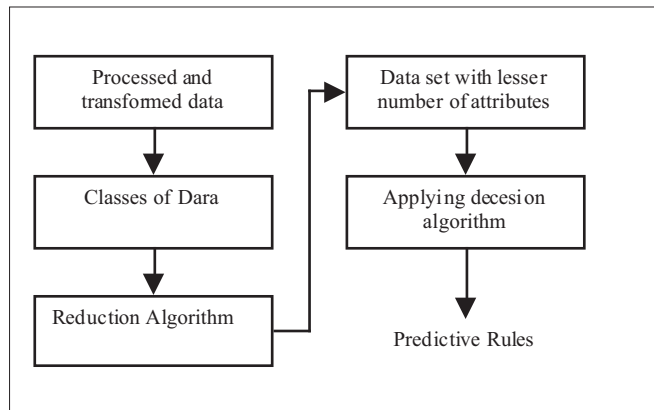
Rough set theory was proposed as a new approach to vague concept description from incomplete data. The rough set theory is one of the most useful techniques in many real life applications such as medicine, pharmacology, engineering, banking and market analysis. This theory provides a powerful foundation to reveal and discover important structures in data and to classify complex objects. One of the main advantages of rough set theory is that it does not need any preliminary or additional information about data. Information about rough sets software for data analysis was given in [4].

Since 1980's, due to the rapid growing hospital information systems, electronic patient records are stored as huge databases at many hospitals. One of the most important problems is that the rules induced from each database (hospital) may be different from those induced from other databases, which are very difficult even for medical experts to interpret. Shusaka Tsumoto introduced rough sets based analysis in order to solve this problem

In our research group at Ain Shams, a rough set-based medical system for mining patient data for predictive rules to determine thrombosis disease was developed in [16] this system aims to search for patterns specific/sensitive to thrombosis disease. This system reduced the number of attri-

butes that describe the thrombosis disease from 60 to 16 significant attribute in addition to extracting some decision rules, through decision applying decision algorithms, which can help young physicians to predict the thrombosis disease. Fig. 2 shows the block diagram of the mining methodology.

**Figure 2.** Rough set mining methodology



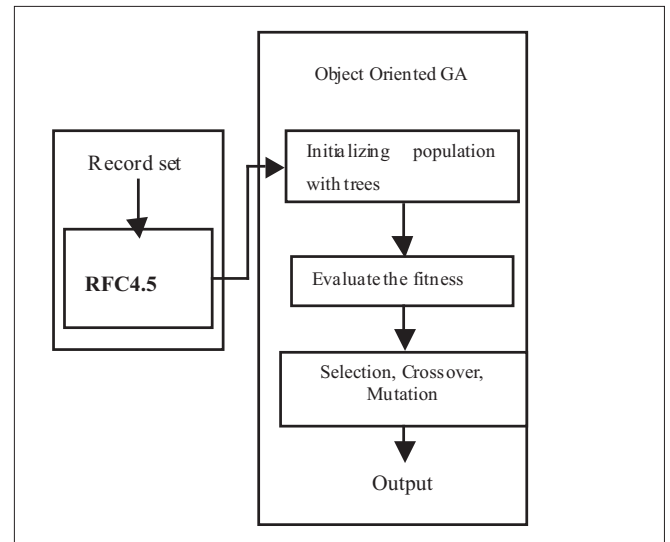
### 5.2 Genetic Algorithms

Classification is an important DM task that has wide of medical diagnosis [5]. Many classifications models have been proposed in the literature such as distributed algorithms, restricted search, data reduction algorithms, parallel algorithms, neural networks and decision trees, genetic algorithms. These approaches either cause loss of accuracy or cannot effectively uncover the data structure. Genetic Algorithms (GA) provide an approach to learning that based loosely on simulated evolution. The GA methodology hinges on a population of potential solutions, and as such exploits the mechanisms of natural selection well known in evolution. Rather than searching from general to specific hypothesis or from simple to complex GA generates successive hypotheses by repeatedly mutating and recombining parts of the best currently known hypotheses. The GA algorithm operates by iteratively updating a poll of hypotheses (population). One each iteration, old members of the population are evaluated according a fitness function. A new generation is then generated by probabilistically selecting the fittest individuals form the current population. Some of these selected individuals are carried forward into the next generation population others are used as the bases for creating new offsprings individuals by applying genetic operations such as crossover and mutation.

In our research group we developed a hybrid classifier that integrates the strengths of genetic algorithms and decision trees. Fig. 3 shows the classifier methodology. The algorithm was applied on a medical database of 20 MB size for predicting

thrombosis disease [7]. The results show that in our classifier is a very promising tool for thrombosis disease prediction in terms of predictive accuracy.

**Figure 3.** GA-based classifier methodology for predicting thrombosis disease.



### 6. Conclusion

This paper describes some of AI-based technologies which are used in developing medical tutoring/learning at Ain Shams University, Cairo, Egypt. The usage of expert systems, neural networks and data mining technologies in developing some of medical prototype systems are explained. A rough set and genetic algorithms based medical systems is used for mining patient data for predictive rules to determine thrombosis disease.

### References

- [1] Lu, R., Cao, C. Chen, Y., and Han, Z. "On Automatic Generation of Intelligent Tutoring Systems", Proceedings of the Int. Conf. Of Artificial Intelligence in Education, Washington D.C., 1995, USA, 67-74.
- [2] Kane, B. and Rucker D. W. AI in medicine. 1998.
- [3] Greer, J. Proceedings of AI-ED 95, World Conference on Artificial Intelligence in Education, Assocaition for Advancement of Computing in Education (AACE), 1995.
- [4] Cios K. J., Pedrycz, W. and Swiniarski, R.W. Data Mining Methods for Knowledge Discovery. Kluwer, 1998.
- [5] Michell, T.M., Machine Learning, McGRAW-HILL, 1997.
- [6] A. M. salem, safia A. Mahmoud., "Mining patient Data Based on Rough Set Theory to Determine Thrombosis Disease", Proceedings of First Intelligence conference on Intelligent Computing and Information Systems, pp 291-296. ICICIS 2002, Cairo, Egypt, June 24-26,2002.
- [7] Abdel-Badeeh M.Salem and Abeer M.Mahmoud, "A Hybrid Genetic Algorithm-Decision Tree Classifier", Proceedings of the 3rd International Conference on New Trends in Intelligent Information Processing and Web Mining, Zakopane, Poland, pp. 221-232, June 2-5, 2003.