

Knowledge Discovery Technique for Web-Based Diabetes Educational System

Satish Kumar David^{1*}, Niraj Upadhayaya², Siddiqui MK³ and Adnan Mahmood Usmani⁴

¹IT Department, Strategic Center for Diabetes Research, King Saud University, Riyadh, Saudi Arabia

²Department of Computer Science & Engineering, Sri K.S.Raju Institute of Technology & Science, Hyderabad, India

³Department of Basic Sciences, Strategic Center for Diabetes Research, King Saud University, Riyadh, Saudi Arabia

⁴Scientific Office, Strategic Center for Diabetes Research, King Saud University, Riyadh, Saudi Arabia

Abstract

Diabetes Mellitus (DM) is the most common metabolic disorder. In developing countries, the prevalence of diabetes is increasing. It is important that every country should assess the magnitude of the problem and take steps to control and prevent DM and provide appropriate care. Despite all the advances in diabetes treatment, education remains the cornerstone of diabetes management. Diabetes education classes are useful for providing general information to be tailored to the specific needs of each patient. The web-based method seems to be effective in the continuing education. Therefore, the web-based method is recommended, as complementary to the face-to-face method, for designing and delivering some topics of continuing education programs. Web-based educational technologies allow diabetes educators to study how diabetes patients learn and which learning strategies are most effective. Since web-based educational systems are capable of collecting huge amounts of diabetic patient profile data, knowledge discovery techniques can be applied to find interesting relationships between attributes of diabetic patient, assessments, and the solution strategies adopted by the diabetic patients. This article presents an approach for predicting diabetic patient performance. The importance of this task lies in improving diabetic patient performance and the effective design of the online diabetes related courses.

Keywords: Knowledge discovery; Patient performance; Web-based educational system; Diabetes Mellitus; Prediction

Introduction

Diabetes Mellitus (DM) is the most common metabolic disorder and its prevalence varies worldwide. In developing countries, the prevalence of diabetes is increasing, where there are, as estimated by the World Health Organization (WHO), around 70 million people suffering from diabetes mellitus [1]. Thus, it is essential that every country attempts to assess the magnitude of the problem and takes steps to control and prevent diabetes mellitus and provide appropriate care.

Despite all the advances in diabetes treatment, education remains the cornerstone of diabetes management. Diabetes education is important in improving diabetes self-management and providing effective diabetes treatment. People with diabetes, unlike those with many other medical problems, can't just take pills or insulin in the morning, and then forget about their health the rest of the day. Differences in diet, exercise levels, stress and other factors may all affect blood glucose levels. So the more people with diabetes learn how these factors affect them, the better control they will be able to achieve. People also need to know what they can do to help prevent or decrease the risk of complications of diabetes. For example, it is estimated that proper foot care can eliminate 75 percent of all amputations performed on people with diabetes. Although diabetes education classes are useful for providing general information, we believe education should be tailored to the specific needs of each patient. An individualized treatment plan can then be developed to address each person's physical, emotional, dietary and educational needs. The web-based method seems to be as effective as the face-to-face method in the continuing education. Therefore, the web-based method is recommended, as complementary to the face-to-face method, for designing and delivering topics of continuing education programs.

The ever-increasing development of network-distributed computing and particularly the rapid expansion of the web have had a broad impact on society in a relatively short period of time. Education for

diabetic patient is on the threshold of a new era based on these changes. Online delivery of educational instruction provides the opportunity to diabetic patient. Many leading medical educational institutions are working to establish an online teaching and learning presence. Several different approaches have been developed to deliver online diabetes education in an academic setting. Michigan State University (MSU) has pioneered some of these systems which provide an infrastructure for online instruction (Multi-Media Physics; CAPA; LectureOnline; PhysNet; Kortemeyer and Bauer, 1999; Kashy et al., 1997, LON-CAPA). This study focuses on the knowledge discovery aspects of the online educational system the Learning Online Network with Computer-Assisted Personalized Approach (LON-CAPA). Also an approach to classifying student characteristics in order to predict performance on assessments based on features extracted from logged data in a web-based educational system.

System

In LON-CAPA, deals with two kinds of large data sets: 1) Educational resources such as web pages, demonstrations, simulations, and individualized problems designed for use on homework assignments, quizzes, and examinations; and 2) Information about users who create, modify, assess, or use these resources. In other words,

***Corresponding authors:** Satish Kumar David, Lecturer, IT Department, Strategic Center for Diabetes Research, College of Medicine, King Saud University, PO Box 245, Riyadh 11411, Saudi Arabia. Tel: +96614724179 ext. 3707; Fax: +96614725682; E-mail: satishkdavid@gmail.com/skumar@ksu.edu.sa

Received November 30, 2010; **Accepted** December 28, 2010; **Published** December 30, 2010

Citation: David SK, Upadhayaya N, Siddiqui MK, Usmani AM (2010) Knowledge Discovery Technique for Web-Based Diabetes Educational System. J Health Med Inform 1:102. doi:10.4172/2157-7420.1000102

Copyright: © 2010 David SK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

we have two ever-growing pools of data. As the resource pool grows, the information from diabetic patients who have multiple transactions with these resources also increases. The LON-CAPA system logs any access to these resources as well as the sequence and frequency of access in relation to the successful completion of any assignment.

The web browser represents a remarkable enabling tool to get information to and from diabetic patients. That information can be textual and illustrated, not unlike that presented in a textbook, but also includes various simulations representing a modeling of phenomena, essentially experiments on the computer. Its greatest use however is in transmitting information as to the correct or incorrect solutions of various assigned exercises and problems. It also transmits guidance or hints related to the material, sometimes also to the particular submission by a diabetic patient, and provides the means of communication with fellow diabetic patients and diabetic educators.

Knowledge discovery

The amount of data stored in databases is increasing at a tremendous speed. This gives rise to a need for new techniques and tools to aid humans in automatically and intelligently analyzing huge data sets to gather useful information. This growing need gives birth to a new research field called Knowledge Discovery in Databases (KDD) or Data Mining, which has attracted attention from researchers in many different fields including database design, statistics, pattern recognition, machine learning, and data visualization.

Our motivation in this study is gaining the best technique for extracting useful information from large amounts of data in an online diabetes educational system, in general, and from the LON-CAPA system, in particular. The goals for this article are: to obtain an optimal predictive model for diabetic patients within such systems, help diabetic patients use the learning resources better, based on the usage of the resource by other diabetic patients in their groups, help diabetic educator design their curricula more effectively, and provide the information that can be usefully applied by diabetic educator to increase diabetic patients learning.

Data Mining is the process of analyzing data from different perspectives and summarizing the results as useful information. It has been defined as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Frawley et al., 1992; Fayyad et al., 1996).

“KDD refers to the overall process of discovering knowledge from data while data mining refers to application of algorithms for extracting patterns from data without the additional steps of the KDD process.” (Fayyad et al., 1996)

The objective of data mining is both prediction and description. That is, to predict unknown or future values of the attributes of interest using other attributes in the databases, while describing the data in a manner understandable and interpretable to humans. Predicting the sale amounts of a new product based on advertising expenditure, or predicting wind velocities as a function of temperature, humidity, air pressure, etc., are examples of tasks with a predictive goal in data mining. Describing the different terrain groupings that emerge in a sampling of satellite imagery is an example of a descriptive goal for a data mining task. The relative importance of description and prediction can vary between different applications. These two goals can be fulfilled by any of a number of data mining tasks including: classification, regression, clustering, summarization, dependency modeling, and deviation detection. (Harrell and Frank, 2001; Montgomery et al., 2001)

Predictive tasks

The following are general tasks that serve predictive data mining goals:

- Classification – to segregate items into several predefined classes. Given a collection of training samples, this type of task can be designed to find a model for class attributes as a function of the values of other attributes (Duda et al., 2001).
- Regression – to predict a value of a given continuously valued variable based on the values of other variables, assuming either a linear or nonlinear model of dependency. These tasks are studied in statistics and neural network fields (Montgomery et al., 2001).
- Deviation Detection – to discover the most significant changes in data from previously measured or normative values (Arning et al., 1996; Fayyad et al., 1996). Explicit information outside the data, like integrity constraints or predefined patterns, is used for deviation detection. Arning et al., (1996) approached the problem from the inside of the data, using the implicit redundancy.

Descriptive tasks

- Clustering – to identify a set of categories, or clusters, that describe the data (Jain & Dubes, 1988).
- Summarization – to find a concise description for a subset of data. There are more sophisticated techniques for summarization and they are usually applied to facilitate automated report generation and interactive data analysis (Fayyad et al., 1996).
- Dependency modeling – to find a model that describes significant dependencies between variables. For example, probabilistic dependency networks use conditional independence to specify the structural level of the model and probabilities or correlation to specify the strengths (quantitative level) of dependencies (Heckerman, 1996).

Discussions

Several Online Education systems [2] such as Blackboard, JUSOOR, WebDiamen, POEM, Virtual University (VU), and some other similar systems have been developed to focus on course management issues. The objectives of these systems are to present courses and instructional programs through the web and other technologically enhanced media. These new technologies make it possible to offer instruction without the limitations of time and place found in traditional university programs. However, these systems tend to use existing materials and present them as a static package via the Internet. There is another approach, pursued in LON-CAPA, to construct more-or-less new courses using newer network technology. In this model of content creation, diabetic educators, and diabetic patients interested in collaboration can access a database of hypermedia software modules that can be linked and combined (Kortemeyer and Bauer, 1999). The LON-CAPA system is the primary focus.

LON-CAPA is a distributed instructional management system, which provides diabetic patients with personalized problem sets, quizzes, and exams. Personalized (or individualized) homework means that each diabetic patient sees a slightly different computer generated problem. LON-CAPA provides diabetic patients and diabetic educator with immediate feedback on conceptual understanding and correctness of solutions. It also provides faculty the ability to augment their courses with individualized, relevant exercises, and develop and share modular online resources. LON-CAPA aims to put this functionality

on a homogeneously distributed platform for creating, sharing, and delivering course content with emphasis on cross-institutional collaboration and intellectual property rights management.

Figure 1.2 shows an overview of this network. All machines in the network are connected with each other through two-way persistent TCP/IP connections. The network has two classes of servers: library servers and access servers. A library server can act as a home server that stores all personal records of users, and is responsible for the initial authentication of users when a session is opened on any server in the network. For authors, it also hosts their construction area and the authoritative copy of every resource that has been published by that author. An Access Server is a machine that hosts diabetic patient sessions. Library servers can be used as backups to host sessions when all access servers in the network are overloaded.

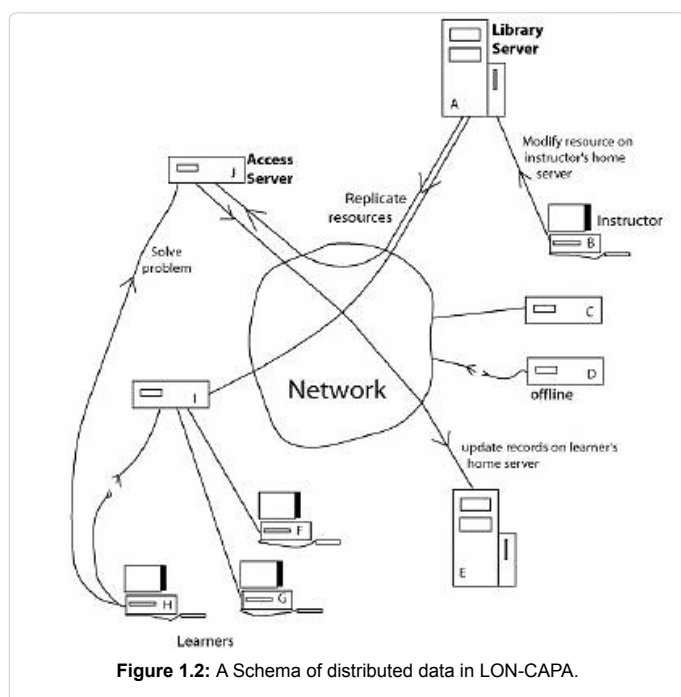


Figure 1.2: A Schema of distributed data in LON-CAPA.

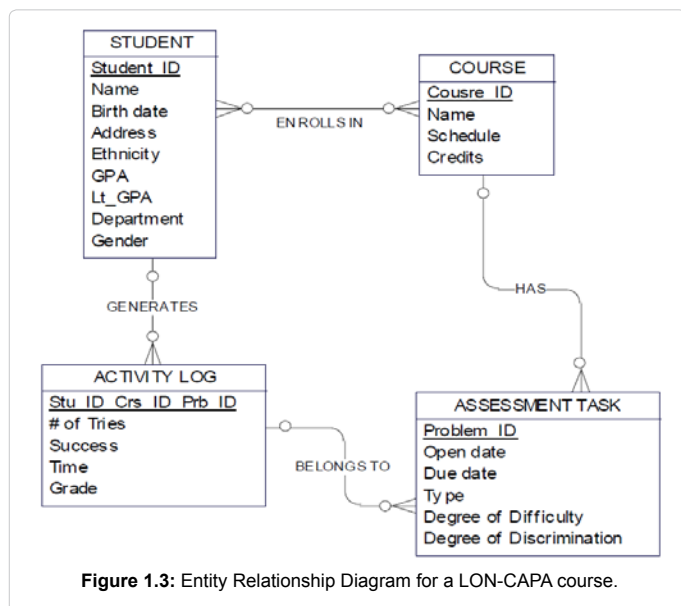


Figure 1.3: Entity Relationship Diagram for a LON-CAPA course.

Educational objects in LON-CAPA range from simple paragraphs of text, movies, and applets, to individualized homework problems. LON-CAPA will allow groups of organizations (departments, universities, schools, commercial businesses) to link their online instructional resources in a common marketplace, thus creating an online economy for instructional resources (lon-capa.org). Internally, all resources are identified primarily by their URL.

LON-CAPA provides three types of resources for organizing a course. LON-CAPA refers to these resources as Content Pages, Problems, and Maps. Maps may be either of two types: Sequences or Pages. LON-CAPA resources may be used to build the outline, or structure, for the presentation of the course to the diabetic patients.

- A Content Page displays course content. It is essentially a conventional html page. These resources use the extension “.html”.
- A Problem resource represents problems for the diabetic patients to solve, with answers stored in the system. These resources are stored in files that must use the extension “.problem”.
- A Page is a type of Map which is used to join other resources together into one HTML page. For example, a page of problems will appear as a problem set. These resources are stored in files that must use the extension “.page”.
- A Sequence is a type of Map, which is used to link other resources together. Sequences are stored in files that must use the extension “.sequence”. Sequences can contain other sequences and pages.

Authors create these resources and publish them in library servers. Then, instructors use these resources in online courses. The LON-CAPA system logs any access to these resources as well as the sequence and frequency of access in relation to the successful completion of any assignment. All these accesses are logged.

One of the most challenging aspects of the system is to provide diabetic educator with information concerning the quality and effectiveness of the various materials in the resource pool on diabetic patient understanding of concepts. These materials can include web pages, demonstrations, simulations, and individualized problems designed for use on homework assignments, quizzes, and examinations. To evaluate resource pool materials, a standardized format is required so that materials from different sources can be compared. This helps resource users to select the most effective materials available.

Predicting diabetic patient learning performance

The objective is to predict the diabetic patients final grades based on the features which are extracted from their (and others) homework data. We design, implement, and evaluate a series of pattern classifiers with various parameters in order to compare their performance in a real data set from the LON-CAPA system. This experiment provides an opportunity to study how pattern recognition and classification theory could be put into practice based on the logged data in LON-CAPA.

Conclusion

In this paper we proposed an approach for predicting diabetic patient performance. This approach can help diabetes educator to design courses more effectively, detect anomalies, inspire and direct further research, and help diabetic patient’s user resources more efficiently. This approach is easily adaptable to different types of courses, different population sizes, and allows for different features to be analyzed. This approach is very useful in identifying those diabetic patients who are at risk of failure, especially in very large classes. This will help the instructor provide appropriate advising in a timely manner.

References

1. The world health report (1997) Conquering suffering, enriching humanity, World Health Organization, Geneva.
2. <http://www.edutools.info>
3. King H, Rewes M (1993) Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. WHO Ad Hoc Diabetes Reporting Group. *Diabetes care* 16:157-177.
4. Mann JI, Pyorala K, Teuscher A (1983) Diabetes in epidemiological perspective. Edinburgh, Churchill Livingstone 122-164.
5. Zimmet P (1983) Epidemiology of diabetes mellitus. In: Ellenberg M, Riffin H, eds. *Diabetes mellitus—therapy and clinical practice*. (3rd edn.) Medical Examination Publishing, New York.
6. Zimmet P, King H, Taylor R, Raper LR, Balkau B et al. (1984) The high prevalence of diabetes mellitus, impaired glucose tolerance and diabetic retinopathy in Nauru —the 1982 survey. *Diabetes research* 1:13-18.
7. Ekoe JM (1986) Recent trends in prevalence and incidence of diabetes mellitus syndrome in the world. *Diabetes research and clinical practice* 1: 249-264.
8. (1979) Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National Diabetes Data Group. *Diabetes* 28: 1039-1057.
9. Fatani HH, Mira SA, El-Zubier AG. (1987) Prevalence of diabetes mellitus in rural Saudi Arabia. *Diabetes care* 10: 180-183.
10. El-Hazmi MAF, Warsy AS (1989) A comparative study of hyperglycaemia in different regions of Saudi Arabia. *Annals of Saudi medicine* 9: 435-438.
11. El-Hazmi MAF et al. (1995) Prevalence of diabetes mellitus in Saudi Arabia. *Saudi medical journal* 16: 294-299.