Available online at www.sciencedirect.com

ScienceDirect

Journal of Electrical Systems and Information Technology 1 (2014) 10–16

www.elsevier.com/locate/jesit

Bayesian based adaptive question generation technique

Nabila Khodeir [a,*], Nayer Wanas [a], Nevin Darwish [b], Nadia Hegazy [a]

[a] *Informatics Department, Electronics Research Institute, Dokki, Giza, Egypt*
[b] *Department of Computer Engineering, Cairo University, Giza, Egypt*

Available online 15 March 2014

Abstract

In this paper we aim to estimate the student knowledge model in a probabilistic domain using automatic adaptively generated assessment questions. The student answers are used to estimate the actual student model. Updating and verification of the model are conducted based on the matching between the student's and model answers. Moreover, a comparative study between using the adaptive and random generated questions for updating the student model is investigated. Results suggest that utilizing adapted generated questions increases the approximation accuracy of the student model by 40% in addition to decreasing of the required assessing questions by 35%.

Keywords: Intelligent Tutoring System; Student modeling; Abduction; Adaptive assessment; Item response theory

1. Introduction

Asking questions is a tool for many systems to achieve a specific goal including assessment and enhancing learners' engagement (Graesser et al., 2005) in Intelligent Tutoring Systems (ITSs). In addition, ITSs rely on assessing of the student answers to presented questions to model his/her knowledge. Based on the student model, ITSs have the ability to personalize their support and interactions for each individual student (Brusilovskiy, 2003).

Automatic generation of questions supports functionality of ITSs, in addition to dialog systems (Pwek, 2010), and Question Answering (QA) systems (Kalady, 2010). Most question generation techniques revolve around linguistic study including syntactic and semantic analysis for the given document to generate questions (Heilman and Smith, 2009; Becker, 2010). In turn, factual and definitional questions are the common types of generated questions in these approaches (Heiman, 2010; Becker, 2010). However, queries associated with some domains cannot be generated or answered based on linguistic analysis. For example, mathematics and physics tutoring systems need to define

* Corresponding author.
E-mail addresses: nabila_khodier@yahoo.com (N. Khodeir), nwanas@ieee.org (N. Wanas), ndarwish@ieee.org (N. Darwish), nhegazy@mcit.gov.eg (N. Hegazy).
Peer review under responsibility of Electronics Research Institute (ERI).

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jesit.2014.03.007

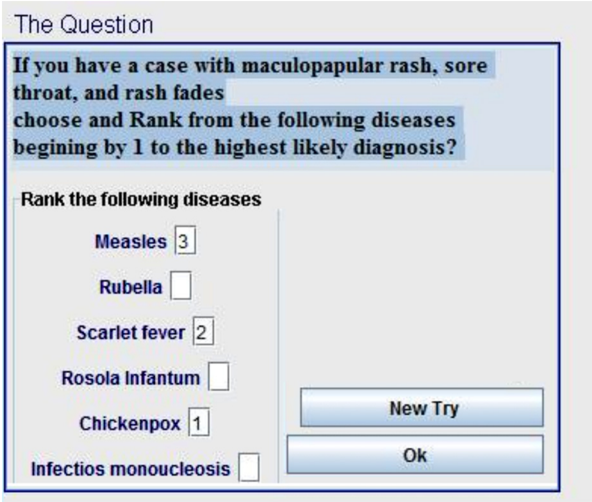

Fig. 1. Question and answer form.

some rules and their applying sequences in solving or generating a specific question. Probabilistic domain represents a difficult problem in this regard. The uncertainty incorporated in the knowledge representation makes generation of questions and answers a difficult task. In such domains, automatic generation of questions and their answers need to be based on a knowledge representation of the domain. In this paper we propose an approach to generate questions and their answers automatically by utilizing Bayesian Network (BN) knowledge representation (Korb and Nicholson, 2011) for probabilistic domains. The purpose of the generated questions is to model the student knowledge within ITS.

Probabilistic domains are domains that consider uncertainty in defining relations between their items. For example, medical domains need to define the probability of association of a symptom to a specific disease. The modeling using BN of such domains is based on defining a set of nodes and a set of directed arcs or links. The nodes in the BN represent a set of random variables $X = x_1, \ldots, x_i, \ldots, x_n$ from the domain. The set of links connects pairs of nodes, $x_i \rightarrow x_j$ representing direct dependencies between variables. The strength of these relations is defined using the Conditional Probability Distribution (CPD) associated with each node. The CPD lists the probability that the child node takes on each of its different values for each combination of values of its parents (Korb and Nicholson, 2011). On the other hand, reasoning in BNs is a process of inferring new information conditioned by observing values of some variables. The process of inference is performed via a flow of information through the network to compute the posterior probability distribution for a set of query nodes given values for some evidence (or observation) nodes.

Probabilistic domains are usually associated by diagnostic questions which require identifying the most probable explanation given a set of evidences. We consider such questions especially in relation to ambiguous cases, where more than one hypothesis that explains the question evidences exist. In such cases the student is asked to provide a ranked list of possible hypotheses for the question evidences. Diagnostic questions for ambiguous cases are chosen since answers for such questions reveal more information about the student knowledge. Consequently fewer questions will be sufficient for the student knowledge modeling.

The generated question takes the following form “If you have a case with evidence_1, evidence_2, evidence_n. Choose and rank from the following hypothesis: Hypothesis_1, Hypothesis_2, Hypothesis_3, . . ., Hypothesis_n.”

The answer form is a ranked list of likely diagnosis hypotheses. Hypotheses are chosen from the available choices that are associated with the question. The student is asked to type the rank value corresponding to the chosen hypothesis. Fig. 1 is a snapshot of the question and answer form.

The rest of this paper is organized as follows. Section 2 presents the proposed question and answer generation techniques. Thereafter, we explore the experimental results that illustrate achieving of generated questions to their purpose of modeling the student knowledge in Section 3. Section 4 presents a discussion and conclusion of the paper.