Automatic Detection of Concepts from Problem Solving Times

P. Boroš, J. Nižnan, R. Pelánek, and J. Řihák

Faculty of Informatics, Masaryk University Brno
{boros,niznan,xpelanek,thran}@mail.muni.cz

Abstract. Intelligent tutoring systems need to know a mapping between particular problems and general domain concepts. Such mapping can be constructed manually by an expert, but that is time consuming and error prone. Our aim is to detect concepts automatically from problem solving times. We propose and evaluate two approaches: a model of problem solving times with multidimensional skill and an application of spectral clustering. The results show that it is feasible to construct a problem-concept mapping from solely the problem solving times and that the results of the analysis can bring an interesting insight.

1 Introduction

One of the functions of intelligent tutoring systems is to provide students feedback on their skills and to adaptively select suitable problems. To this end, the system needs to understand its target domain, e.g., to have a mapping between particular problems and general domain concepts (e.g., fractions, linear functions, trigonometry). Such mapping can be constructed manually by an expert, but that is time consuming and error prone. Thus it is desirable to construct and validate such mappings automatically.

In this work we study techniques to automatically determine concepts from problem solving times. To this end, we propose and evaluate two approaches. The first approach is a multidimensional extension of the model of problem solving times, which we introduced in previous work [3]. This model is analogous to the Q-matrix approach [1], the main difference is that a standard Q-matrix is used with discrete values (0/1), whereas in our setting we use continuous values. The second approach is to determine the concepts by a clustering technique, particularly by the spectral clustering method [4]. For evaluation we use data from the Problem Solving Tutor [3].

The results show that it is feasible to automatically detect concepts (i.e., similarity between problems) from problem solving times. We present a specific example which shows an interesting and useful output of the automatic detection of concepts. The two studied methods, although using completely different approaches, give similar overall results. The approach based on multidimensional model is less stable than spectral clustering, but can be naturally used for validation and improvement of a Q-matrix provided by an expert.

2 Techniques for Detection of Concepts

In the following we assume that we have a set of students, a set of problems, and data about problem solving times: t_{sp} is a logarithm of time it took a student s to solve a problem p. The first approach is based on a previously described model of problem solving times [3]. The basic structure of the model is simple – it assumes a linear relationship between the logarithm of a time and a skill: $t_{sp} = b_p + q_p\theta_s + \epsilon$, where t_{sp} is the logarithm of a problem solving time for a student s and a problem p, b_p is the basic difficulty of a problem p, q_p is a discrimination factor of a problem p, θ_s is a skill of a student s, and ϵ is Gaussian noise. A more detailed discussion of the model is given in [3].

Here we consider an extension of this model with k dimensional skills: $t_{sp} = b_p + \boldsymbol{q_p}^T\boldsymbol{\theta_s} + \epsilon$, where $\boldsymbol{q_p}$ is a k dimensional discrimination vector and $\boldsymbol{\theta_s}$ is a k dimensional skill vector. The model is analogical to widely used Q-matrix models [1]. The main difference is that standard Q-matrices are typically used in the setting of test questions with binary response ($0 -$ incorrect, $1 -$ correct), with the Q-matrix entries and student skills being also binary (the model specifies probability of a correct answer using noisy and/or function).

In our setting it is natural to allow both Q-matrix values (discrimination factors) and skills to be continuous. The estimation of the parameters can be done by stochastic gradient descent, analogically to the model with one dimensional skill [3]. The main complication with respect to the one dimensional case is a suitable initialization of the gradient descent.

The second approach is spectral clustering, which is a popular clustering technique based on linear algebra [4]. The main principle of the algorithm is the following. At first, a similarity graph for the data is created; construction of this graph is specific to a domain of application. At second, a Laplacian matrix of the similarity graph is constructed and its first n eigenvectors are computed. At third, the eigenvectors are used to transform original data into points in R^n; these points are clustered using the standard k-means algorithm (an illustration is provided in Fig. 1.).

Note that only the first step is problem specific, the other two steps are generic. In our case for each pair of problems we define their similarity as a Spearman's correlation coefficient of times of shared students (those who solved both problems). Based on the computed similarity matrix, we construct k-nearest-neighbours graph (connecting each node with k most similar nodes); where k is one half of the number of problems. For our data the spectral clustering method is quite stable and not susceptible to details (e.g., a choice of k or an exact version of the algorithm).

3 Evaluation

We evaluated the described techniques over data from the Problem Solving Tutor (`tutor.fi.muni.cz`) [3]. To evaluate the identification of concepts we performed the following experiment. We mix data from the Problem Solving Tutor for two different types of problems and remove information about the type of the problem from the data. Then we let an algorithm analyze the data and cluster problems into two groups. The performance is measured by the number of correctly clustered problems (reported below as percentage of all problems). For the experiment we used 8 most solved problem types from the Problem Solving Tutor. The experiment was performed on all pairs of these problem types. For each pair we consider only data about students who solved at least 10 problems from both problem types. On average the data for each problem pair contain 150 problem instances and 150 students.

The overall mean performance of spectral clustering is 86.5%, the model with two skills achieves very similar overall results (85.9%). We also evaluated clustering using only the standard k-means algorithm (each problem is represented as a vector of correlation coefficients with other problems), the performance in this case was slightly worse (83.5%). All of these techniques are partially stochastic, so we measured the performance over multiple runs. Spectral clustering yields very consistent results, whereas the model based classification (which uses stochastic gradient descent with randomized initialization) has in few cases large variance.

The performance differs for individual problems. Results are very good (over 90%, up to 99%) for problems which strongly depend on logical reasoning skills and thus the noise in the data is low. On the other hand, results are poor for a geometric puzzle, where the noise in data is quite high since the puzzle is based more on insight and luck than on skill.

We analyzed one of the problems in more detail – a Binary crossword problem. The goal is to fill a grid with zeros and ones in such a way that all specified conditions are met (see Fig. 1.). This setting can be used for easy problems for practicing basics of binary numbers (a) and logic operations (b), but also for more challenging problems where the specified conditions are given in self-referential crossword manner, which leads to quite entertaining practice of binary numbers and logic operations (c, d).

There are 55 instances of Binary crosswords and they can be naturally divided to three main groups: examples based on knowledge of binary notation, examples which use logical operations, and the self-referential crossword examples which usually combine different types of conditions and require deeper thinking. The Problem Solving Tutor contains manually created classification of examples into these three types (binary numbers, logic operations, and crosswords). We used the spectral clustering method to compute 3 clusters and compared the results with the manual labeling of examples – the agreement of the classifications was about 80%. Most differences can be intuitively explained, and in fact in most cases the misclassification brings an useful insight, often showing inappropriateness of the manual labeling.

One of the advantages of spectral clustering is the possibility to use the computed eigenvectors to plot data in a low dimensional space. In Fig. 1. we can see that instances a, b, c, which are typical examples of their groups, are placed in the middle of their clusters. Instance d has a form of a self-referential crossword but strongly uses the concept of logic operations and the location of this example

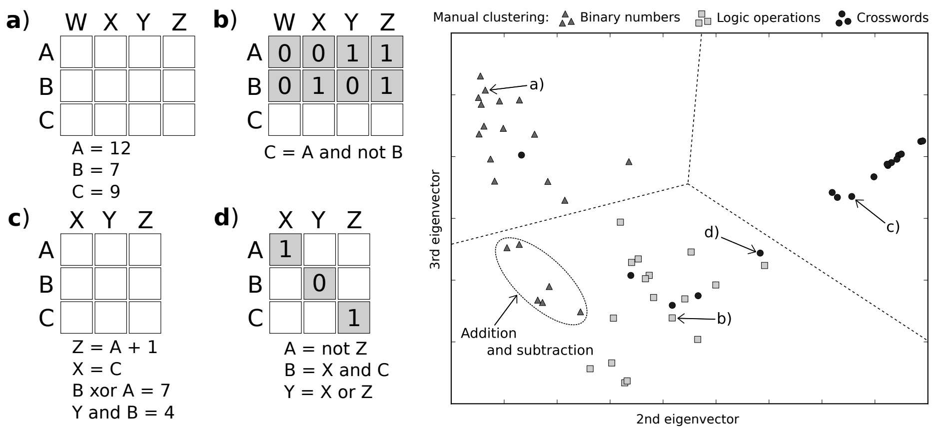

Fig. 1. Examples of Binary crossword problems and projection of all problems onto a plane by spectral clustering (with algorithmically determined clusters)

corresponds to this observation. Another similar example is a circle in Binary number cluster which also has a form of crosswords, but solving requires only ability to write binary numbers. Interesting results were obtained for examples based on addition and subtraction of binary numbers. These examples were manually labelled as "binary numbers", but Fig. 1. suggests that these examples are slightly different then other binary numbers examples and are closer to the "logic operations" cluster.

This kind of analysis can be done also using the model with multiple skills. In this case it is natural to initialize the model according to the provided labeling, fit the model to data, and then check which problems deviate most from the initialization (thus performing a Q-matrix validation [2]). We have performed this analysis for the Binary crossword problem, the results are similar to the above presented results obtained through spectral clustering.

References

1. T. Barnes. The q-matrix method: Mining student response data for knowledge. In *American Association for Artificial Intelligence 2005 Educational Data Mining Workshop*, 2005.
2. J. De La Torre. An empirically based method of q-matrix validation for the dina model: Development and applications. *Journal of educational measurement*, 45(4):343–362, 2008.
3. P. Jarušek and R. Pelánek. Analysis of a simple model of problem solving times. In *Proc. of Intelligent Tutoring Systems (ITS)*, volume 7315 of *LNCS*, pages 379–388. Springer, 2012.
4. A.Y. Ng, M.I. Jordan, Y. Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.