Using Problem Solving Times and Expert Opinion to Detect Skills

Juraj Nižnan
Masaryk University Brno
niznan@mail.muni.cz

Radek Pelánek
Masaryk University Brno
xpelanek@fi.muni.cz

Jiří Řihák
Masaryk University Brno
thran@mail.muni.cz

ABSTRACT

Construction of a mapping between educational content and skills is an important part of development of adaptive educational systems. This task is difficult, requires a domain expert, and any mistakes in the mapping may hinder the potential of an educational system. In this work we study techniques for improving a problem-skill mapping constructed by a domain expert using student data, particularly problem solving times. We describe and compare different techniques for the task – a multidimensional model of problem solving times and supervised classification techniques. In the evaluation we focus on surveying situations where the combination of expert opinion with student data is most useful.

1. INTRODUCTION

One of important aspects of development of adaptive educational systems is the construction of a mapping between educational content (questions, problems) and latent skills (also denoted as knowledge components or concepts). This mapping is important for student skill estimation, which guides the adaptive behaviour of systems, and is typically constructed by a human and since it is a difficult process, it requires a domain expert. The labeling of items, particularly for large item pools, may be time-consuming, and consequently the process is rather expensive. Another approach is to use automatic construction of the mapping from the data (e.g. Q-matrix method [1, 2]). To be reliable, the automatic approach needs large amount of data. Synergy of these two approaches (e.g. [4]) may bring useful results. We can use a human expert to provide initial labeling of problems and then automatic methods can be used to detect errors that the human might have introduced and to fix them.

Depending on the quality of the provided expert labeling and amount of data, there are three possible scenarios. If the number of expert errors is small or the data are insufficient, it is best to use just the expert opinion (donated as E-zone). If the expert makes lot of mistakes and large data are available, then it is best to use just the data (D-zone). We are interested in the region between these two cases, when it is most advantageous to combine both the expert input and available data (ED-zone). Our aim is to explore techniques for such combination and to map the size of this region.

2. TECHNIQUES

In the following we assume that we have a set of students S, a set of problems P, and data about problem solving times: $t_{s,p}$ is a logarithm of time it took a student $s \in S$ to solve a problem $p \in P$. We have an expert labeling $l_E : P \to \Sigma$ where Σ is the set of skills. The expert labeling may contain some mistakes when compared to a correct hidden labeling l. The output of our algorithms is some other labeling l_A that may be different from l_E. The goal of our algorithms is to provide a more accurate labeling (according to l) than l_E.

2.1 Model with Multidimensional Skill

In this section we introduce a extension of model described in [5] for predicting how much time it takes a student to solve a particular problem. The model uses a few latent attributes: problem difficulty b_p, student skill β_s, problem skill vector $\boldsymbol{q}_p$ and a student skill vector $\boldsymbol{\theta}_s$. It assumes the following relationship between the attributes: $t_{s,p} = b_p\beta_s + \boldsymbol{q}_p^\intercal\boldsymbol{\theta}_s + \epsilon$. The vector $\boldsymbol{q}_p$ represents the weight of individual skills in the problem p. The vector $\boldsymbol{\theta}_s$ can be interpreted as the values of skills the student s has.

This model is supervised in a sense that it is learning to predict the student solving times. As a byproduct we get the Q-matrix $\boldsymbol{Q}$ which represents the problem-skill mapping that we are interested in. The objective of the model is to minimize the squared prediction error. To get the values of the parameters we use stochastic gradient descent with initial Q-matrix provided by expert labeling. After the algorithm terminates we can check for discrepancies between the expert Q-matrix and the Q-matrix outputted by the parameter estimation algorithm. We will assume that these discrepancies are expert mistakes.

2.2 Supervised Learning

The main idea of using supervised classification methods can be illustrated by the most straightforward approach which uses k-NN (k-nearest neighbors) algorithm and Spearman's correlations $r(p_i, p_j)$ of problems p_i, p_j as a measure of problem similarity. We assume that the most correlated problems belong to the same skill and thus have the same labels. So for problem p_i a new label $l_A(p_i)$ will be the most common label (provided by expert) among the k most correlated problems from P with problem p_i. This approach can find some mistakes, however it brings only small improvement of expert labeling l_E.

Similarly we can use different classification methods with different metric. A problem p_i can be represented as a vector $\boldsymbol{r_{p_i}} = \{r(p_i, p_j)\}_{1 \leq j \leq |P|}$ and Euclidean distance of these problems can measure similarity of problems (we assume that two similar problems have similar correlations with other problems). As classifier we have chosen logistic regression, which is more sophisticated but still computationally fast.

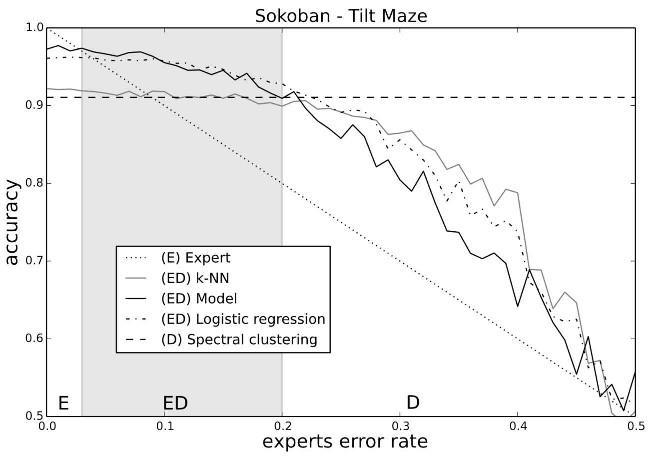

Figure 1: Comparison of techniques for particular situation. The ED-zone is marked for the model.

3. EVALUATION

3.1 Data and Experiment Setup

To evaluate our algorithms we used real data from a Problem Solving Tutor [6]. It is a free web-based tutoring system for practicing problem solving; it is available at `tutor.fi.muni.cz`. To simulate multiple skills for the evaluation purposes we mixed data from k problems together. Each problem type represents a single skill (or label). An expert is simulated by taking the correct labeling and introducing some random mistakes with rate $p_e \in [0, 0.5]$. Hence in this situation (as opposed to standard setting), we know the correct "latent" skills and thus we can measure accuracy of a method as the portion of the final labels assigned correctly. The expected accuracy of an expert (E) is $1 - p_e$. Spectral clustering method (see [3]) was used for the evaluation of the D approach. Finally the expert labeling was used in the ED approaches described in section 2.

3.2 Results

Figure 1 shows the comparison of the accuracies of the E, ED and D approaches. We can denote three zones within expert error rate based on which approach (E, ED or D) performs the best. We are interested particularly in the ED-zone, where the newly introduced approaches are the best, specifically in its position and width, which tells us for which values of p_e these approaches are a good choice.

The figure shows that the algorithm based on k-NN brings only small improvement. The other two approaches are significantly better and to each other comparable, however the algorithm based on logistic regression is significantly faster, because it works only with correlation vectors, which substantially reduces the amount of data. On the other hand approach based on model gives more information about problem-skill mapping, because it provides Q-matrix and not only labeling.

Experiments for other problem combinations showed that the size of the zone grows with decreasing performance of D approach and with number of skills. For larger numbers of skills the zone becomes dominant.

4. DISCUSSION

Our experiments address two types of questions: "how" and "when". The "how" question is concerned with the choice of suitable technique for combining expert opinion and student data. Here the results suggest that on one hand the choice of technique is important – note that two similar supervised approaches (k-NN, logistic regression) achieve quite different results. On the other hand, two significantly different approaches (the multidimensional model and logistic regression) achieve very similar results. The "when" question is concerned with mapping when it is useful to use the combination of expert opinion and student data. The results show that this "zone" is sufficiently large to deserve attention and it is useful to combine the expert opinion with student data for large range of quality of expert input.

5. 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] T. Barnes. Novel derivation and application of skill matrices: The q-matrix method. *Handbook on Educational Data Mining*, 2010.

[3] P. Boroš, J. Nižnan, R. Pelánek, and J. Řihák. Automatic detection of concepts from problem solving times. In *Proc. of International Conference on Artificial Intelligence in Education (AIED 2013)*, volume 7926 of *LNCS*, pages 595–598. Springer, 2013.

[4] M. Desmarais and R. Naceur. A matrix factorization method for mapping items to skills and for enhancing expert-based q-matrices. In H. Lane, K. Yacef, J. Mostow, and P. Pavlik, editors, *Artificial Intelligence in Education*, volume 7926 of *Lecture Notes in Computer Science*, pages 441–450. Springer Berlin Heidelberg, 2013.

[5] 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.

[6] P. Jarušek and R. Pelánek. A web-based problem solving tool for introductory computer science. In *Proc. of Innovation and technology in computer science education*, pages 371–371. ACM, 2012.