Personalized Forecasting Student Performance

Nguyen Thai-Nghe
University of Hildesheim
31141 Hildesheim, Germany
nguyen@ismll.de

Tomáš Horváth
University of Hildesheim
31141 Hildesheim, Germany
horvath@ismll.de

Lars Schmidt-Thieme
University of Hildesheim
31141 Hildesheim, Germany
schmidt-thieme@ismll.de

***Abstract*—This work proposes a novel approach - *personalized forecasting* - to take into account the sequential effect in predicting student performance (PSP). Instead of using all historical data as other methods in PSP, the proposed methods only use the information of the individual students for forecasting his/her own performance. Moreover, these methods also encode the "student effect" (e.g. how good/clever a student is, in performing the tasks) and "task effect" (e.g. how difficult/easy the task is) into the models. Experimental results show that the proposed methods perform nicely and much faster than the other state-of-the-art methods in PSP.**

***Keywords*-Predicting student performance; Personalized forecasting; Sequential effect**

I. Introduction

Educational data mining has been taken into account recently [1], especially in predicting student performance (PSP) [2]. Cen et al. [3] have shown that an improved model for PSP could save millions of hours of students' time and effort. Moreover, many universities are extremely focused on assessment, thus, the pressure on "teaching and learning for examinations" leads to a significant amount of time spending for preparing and taking standardized tests. From an educational data mining point of view, *a good model which accurately predicts student performance could replace some current standardized tests* [2], [4].

To address the PSP, many works have been published but they rely on traditional methods such as logistic regression [3], linear regression [4], and so on [1]. Recently, Thai-Nghe et al. [5] have reported that PSP can be casted as rating prediction in recommender systems, thus, they proposed using matrix factorization to cope with this problem. The authors have shown that using recommendation techniques could improve the prediction results compared to regression methods but they have not taken the sequential effect into account. Obviously, from the educational point of view, the learner's knowledge cumulates and improves over the time, thus, sequential effect is an important information for such prediction tasks. In other domains (e.g. stock-market), to address the sequential effect, forecasting techniques are quite appropriate. As far as we know, forecasting techniques have not been used for PSP, especially for taking into account the sequential effect in PSP.

This work proposes the personalized forecasting methods for PSP. Different from other literature, e.g. [2], [5], which use all historical data to form the prediction models, the proposed approach only uses historical information of individual student to forecast/predict his/her own performance. Moreover, the proposed methods also incorporate the "student effect/bias" (e.g. how good/clever a student is, in performing the tasks) and "task effect/bias (e.g. how difficult/easy the task is) to the models.

II. Predicting Student Performance (PSP)

The problem of PSP is to predict the likely performance of a student for some exercises (or part thereof such as for some particular steps) which we call the *tasks*. The task could be to solve a particular step in a *problem*, to solve a whole problem or to solve problems in a *section* or *unit*, etc. Tasks can be located in a topic hierarchy: $unit \supseteq section \supseteq problem \supseteq step$ (see the article [2] for details). PSP can be casted as rating prediction task in recommender systems since student s, task i and performance p would be *user, item* and *rating*, respectively [5]. This problem can also be *casted as forecasting problem* to take into account the *potentially sequential effects* (e.g. describing how students gain experience over time), as mentioned in [2]. For personalization purpose, we denote p^s and $\hat{p}^s$ as the actual performance and the predicted performance of a given student s, respectively.

III. Methods

Personalized Forecasting: Different from other approaches in PSP, which uses all the historical data to form the models, we only use the historical data of the individual student for forecasting/predicting his/her performance. An example for justifying this choice is: *"Mary is a clever student. Her performance is always better/higher than the other students. Thus, using her performance information to predict/forecast the others would not fit, and we should use one's performance for predicting oneself"*.

Moreover, we use a parameter L to control the length of the historical data (the number of previous steps used for forecasting. We call this method *"hist"*). In another approach, instead of using L, we can use all historical data in the same *unit* and *section* for forecasting the performance of new problems in that unit and section (we call this method *"sec"*). Figure 1 illustrates these two approaches. For examples, to forecast Mary's performance on the steps of "Problem 3", we can use all previous steps with length L including the steps of other previous units and sections, or we can use the previous steps in the same "Unit 5" and "Section 5" (e.g. steps of "Problem 1" and "Problem 2") to form the model.

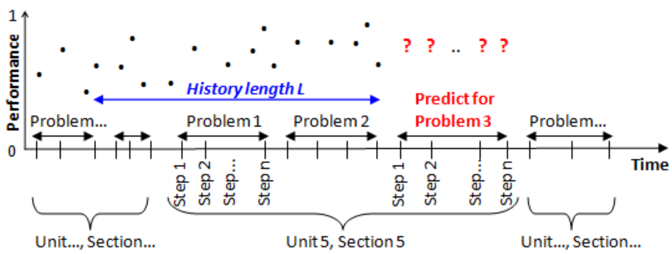

Figure 1. An illustration of personalized forecasting

The personalized forecasting for student s at time $t+1$ using single exponential smoothing [6] is determined by:

$$\hat{p}^s_{t+1} = E^s_t \tag{1}$$

where E^s_t is the smoothing value at time t ($t > 2$):

$$E^s_t = \alpha p^s_{t-1} + (1-\alpha)E^s_{t-1} \quad 0 < \alpha < 1 \tag{2}$$

where p^s_{t-1} and E^s_{t-1} are the actual performance and the smoothing value of student s at time $t-1$, respectively; α is a smoothing constant. Also, we can initialize $E^s_2 = p^s_1$. We call this method **N-PSEF** (Non-biased Personalized Single Exponential smoothing Forecasting).

With double exponential smoothing forecasting, the equation (1) will become:

$$\hat{p}^s_{t+1} = E^s_t + T^s_t \tag{3}$$

where

$$E^s_t = \alpha p^s_{t-1} + (1-\alpha)(E^s_{t-1} + T^s_{t-1}) \quad 0 < \alpha < 1 \tag{4}$$

and $T^s_t = \beta(E^s_t - E^s_{t-1}) + (1-\beta)T^s_{t-1} \quad 0 < \beta < 1$ is the trend value at time t. There is no E^s_1. We can initialize $E^s_2 = p^s_1$ and $T^s_2 = p^s_2 - p^s_1$. We call this method **N-PDEF** (Non-biased Personalized Double Exp. smoothing Forecasting).

Personalized Forecasting with "Student-Task-Bias": We adopt the idea from recommender systems to employ the "user bias" and "item bias" to the models [7]. On the educational setting the user and item bias are, respectively, the student and task biases. They model how good a student is (i.e. how likely is the student to perform a task correctly) and how difficult/easy the task is (i.e. how likely is the task to be performed correctly) [2]. With biases, the prediction function for student s and task i is determined by

$$\hat{p}^{b_{si}} = \mu + b_s + b_i \tag{5}$$

where μ is the global mean (average performance of all students and tasks in the training set $\mathcal{D}^{train}$),

$$\mu = \frac{\sum_{p \in \mathcal{D}^{train}} p}{|\mathcal{D}^{train}|} \tag{6}$$

b_s is student bias (average performance of student s deviated from the global mean):

$$b_s = \frac{\sum_{p^s \in \mathcal{D}^{train}} (p^s - \mu)}{|p^s \in \mathcal{D}^{train}|} \tag{7}$$

and b_i is task bias (average performance on task i deviated from the global mean):

$$b_i = \frac{\sum_{p^i \in \mathcal{D}^{train}} (p^i - \mu)}{|p^i \in \mathcal{D}^{train}|} \tag{8}$$

Using these biases, the forecasting function (1) for student s at time $t+1$ now becomes

$$\hat{p}^s_{t+1} = \alpha \hat{p}^{b_{si}} + E^s_t \tag{9}$$

where E^s_t can be obtained from single (eq. 2) or double (eq. 4) exponential smoothing. We call these methods **B-PSEF** and **B-PDEF**, respectively (Biased Personalized Single/Double Exponential smoothing Forecasting).

Personalized Forecasting with "Discounted-Mean": The true fact is that "memory of human is limited", thus, the students could forget what they have studied in the past, e.g., they could perform better on lessons they have learned recently than on such they have learned last year or before. Moreover, in education point of view: "The more the learners study the better the performance they get" and "their knowledge can cumulate and improve over the time", thus, the sequential effect is an important factor for PSP. Instead of using smoothing values as in the previous sections, we now use a discounted mean value Θ, which reduces the weight controlled by parameter δ when going back to the history:

$$\Theta = \frac{\sum_{t=T-L+1}^{T} p^s_t \cdot e^{t \cdot \delta}}{\sum_{t=T-L+1}^{T} e^{t \cdot \delta}} \qquad 0 < \delta < 1 \tag{10}$$

where T is the current time. The forecasting function (9) now becomes

$$\hat{p}^s_{t+1} = \alpha \hat{p}^b + (1-\alpha)\Theta \tag{11}$$

We call this method **B-PDMF** (Biased Personalized Discounted-Mean Forecasting).

IV. Evaluation

Data sets: Two data sets are collected from the KDD Challenge 2010 (pslcdatashop.web.cmu.edu/KDDCup). The Algebra (Bridge) has 8,918,054 (20,012,498) records for training, and 508,912 (756,386) records for testing. The central element of interaction between the students and the tutoring system is the *problem*. Every problem belongs into a hierarchy of *unit* and *section*, and consists of many individual *steps*. Target of the predicting/forecasting is the *correct* *first attempt* (CFA) information which encodes whether the student successfully completed the given step on the first attempt. The prediction would then encode the *certainty* that the student will succeed on the first try.

Evaluation metric and model setting: Root mean squared error (RMSE) is used to evaluate the models. We initialize the smoothing value by averaging the previous performances, e.g. $E_2^s = \frac{\sum_{t=T-L}^{T} p_t^s}{L-1}$ since preliminary results shown that this initialization gives the results better than initializing with $E_2^s = p_1^s$. Hyperparameter search is used to determine the best hyperparamaters (in term RMSE) for all methods. The forecasting value is bounded with 0/1 in case it exceeds the interval [0..1].

Comparison with other methods: The proposed approaches were compared with: Original single/double exponential forecasting (SEF/DEF) [6], *global mean*, *student mean*, and *student-task bias* (adapted from the user-item-baseline in [7]). We also compare our approach with *matrix factorization* and logistic regression as described in previous works [2], [5].

Empirical results: From the experiments, we found that the *hist* approach works slightly better than the *sec*. This could implicitly mean that to solve the new problems, the learners need all the previous cumulated knowledge rather than only the knowledge in the same unit and section. We use the *hist* strategy for the rest experiments.

Figure 2a compares the root mean squared error (RMSE) of original (non-personalized) forecasting (SEF and DEF) with personalized forecasting methods (N-PSEF and B-PSEF). The personalized methods outperform the non-personalized ones. Moreover, using personalized forecasting and taking into account the "student and task effect" (B-PSEF) significantly improve the results compared to the non-personalized SEF and DEF.

Figure 2b compares the RMSE of personalized forecasting (B-PSEF, B-PDEF, B-PDMF) with the other methods. The proposed methods also outperform the others including the state-of-the-art matrix factorization. [5] reported that linear regression and logistic regression give nearly similar results on these data, so we did not compare with the linear one.

The proposed methods build the models for each student individually, thus, they need not so much computer memory to deal with large datasets while running quite fast. For example, using Bridge, on average B-PSEF only needs **5.3 seconds** (depending on the length L) for both training and testing phases, while logistic regression and matrix factorization need 146.0 and 1299.8 seconds, respectively. In addition, the personalized methods outperform the non-personalized ones. Thus, the results more or less reflect the natural fact that *"The knowledge of human is diverse; thinking and performing of one student differs from another one, so we should not use the performance of someone to predict/forecast for someone else"*. Personalized forecasting

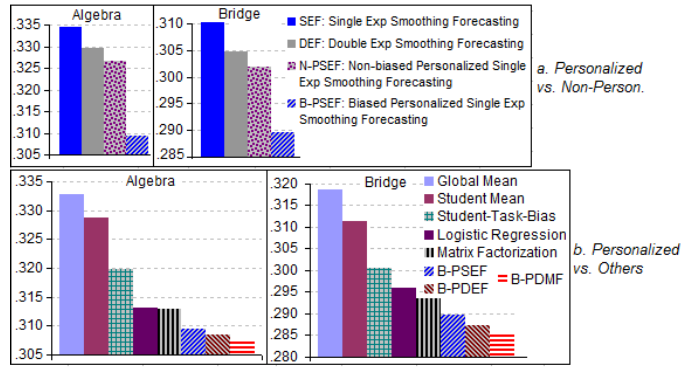

Figure 2. RMSE: Personalized vs. non-personalized forecasting.

approach for PSP shows promising results, especially when *the cumulative knowledge of the learners should be taken into account*.

V. Conclusions

We proposed the personalized forecasting methods, which are very simple but efficient approach, for predicting student performance (PSP). These methods can also take into account the "student effect" and the "task effect" in PSP. The personalized forecasting methods perform nicely and much faster than the other state-of-the-art methods in PSP.

Acknowledgments

The first author was funded by the TRIG project of Cantho university, Vietnam. Tomáš Horváth is also supported by the grant VEGA 1/0131/09.

References

[1] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of Educational Data Mining*. Chapman and Hall/CRC, October 2010.

[2] N. Thai-Nghe, L. Drumond, T. Horvath, A. Krohn-Grimberghe, A. Nanopoulos, and L. Schmidt-Thieme, "Factorization techniques for predicting student performance," in *Educational Rec. Syst. and Tech.: Practices and Challenges (In press)*, O. C. Santos and J. G. Boticario, Eds. IGI Global, 2011.

[3] H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis: A general method for cognitive model evaluation and improvement," in *ITS*. Springer, 2006, pp. 164–175.

[4] M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," *User Modeling and User-Adapted Interaction*, vol. 19, no. 3, pp. 243–266, 2009.

[5] N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme, "Recommender system for predicting student performance," *Procedia Computer Science*, vol. 1, no. 2, pp. 2811 – 2819, 2010.

[6] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Springer, March 2002.

[7] Y. Koren, "Factor in the neighbors: Scalable and accurate collaborative filtering," *ACM Transactions on Knowledge Discovery from Data*, vol. 4, no. 1, pp. 1–24, 2010.