



# Self-paced Graph Memory Network for Student GPA Prediction and Abnormal Student Detection

Yue Yun<sup>1</sup>, Huan Dai<sup>1</sup>, Ruoqi Cao<sup>2</sup>, Yupei Zhang<sup>1</sup>  , and Xuequn Shang<sup>1</sup> 

<sup>1</sup> School of Computer Science, Northwestern Polytechnical University,  
Xi'an 710129, Shaanxi, China

{yundayue,daihuan}@mail.nwpu.edu.cn, {ypzhaang,shang}@nwpu.edu.cn

<sup>2</sup> National Research University Higher School of Economics, 20 Myasnitskaya Ulitsa,  
Moscow 101000, Russia  
rcao@hse.ru

**Abstract.** Student learning performance prediction (SLPP) is a crucial step in high school education. However, traditional methods fail to consider abnormal students. In this study, we organized every student's learning data as a graph to use the schema of graph memory networks (GMNs). To distinguish the students and make GMNs learn robustly, we proposed to train GMNs in an “easy-to-hard” process, leading to self-paced graph memory network (SPGMN). SPGMN chooses the low-difficult samples as a batch to tune the model parameters in each training iteration. This approach not only improves the robustness but also rearranges the student sample from normal to abnormal. The experiment results show that SPGMN achieves a higher prediction accuracy and more robustness in comparison with traditional methods. The resulted student sequence reveals the abnormal student has a different pattern in course selection to normal students.

**Keywords:** Student learning performance prediction · Self-paced learning · Graph memory networks · Abnormal student detection

## 1 Introduction

A crucial demand in higher education is SLPP [1, 2]. Unfortunately, almost all of related researches [3–6] mainly focus on GPA prediction. They treated GPA as the only indicator for evaluating students and failed to consider one abnormal situation: there are some students whose learning patterns are clearly different from others, e.g., promising students or gifted students. Note, they are less but important, and GPA is not enough.

For identifying them, we propose an effective method with strong interpretability in educational sense, and we obtain the inspiration from regression models: students from mentioned abnormal situation are equal to outliers of

training samples in regression models. Then, we aim to design a robust prediction model for the task of SLPP, and identify students in mentioned abnormal situation, i.e., ASD task. In addition, for mining underlying information, we proposed graph memory networks (GMN) model [7] as our basic model. Then, to improve the robustness of GMNs and enhance the precision of training loss, we combine self-paced learning [8] (SPL) with GMNs, and proposed SPGMN algorithm. Specially, our two tasks are combined with SPGMN for better interpretability: we train a SPGMN model for SLPP task on training samples, then, we apply the trained model to handle the ASD task.

**Definition 1.** *Abnormal student is the student whose learning pattern is clearly different from others. And, ASD is the abbreviation of **abnormal student decision**. Specially, this definition is with respect to the data aspect.*

Summarily, the contributions of our work are: 1) SPGMN can enhance the robustness of GMNs model in SLPP task; 2) improve the precision and interpretability of ASD task; 3) Experimental results indicate that SPGMN can really enhance the robustness of GMNs. The ASD experiment proves that GPA is not enough, and we propose learning pattern as an supplement.

## 2 Proposed Methods

**Self-paced Graph Memory Network.** SPL incorporates a self-paced function and a pacing parameter into the learning objective of GMN to optimize the order of samples and model parameters. Each sample is assigned to a weight to reflect the easiness of the sample. Then, these weighted samples are gradually involved into training from easy to complex, as shown as follows:

$$\min_{\mathbf{w}, \mathbf{v}} \mathbb{E}(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^n \mathbf{v}_i \cdot \ell_i(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda^k) \quad (1)$$

where  $y_i$  is the ground-truth label of the  $i$ -th graph,  $\ell(\cdot)$  is the loss function of GMN,  $\mathbf{x}_i$  is the  $i$ -th trained graph,  $\mathbf{w}$  is the parameters,  $f(\mathbf{v}; k)$  is a dynamic self-paced function with respect to  $v \& k$  (e.g.,  $f(v, \lambda^k) = \lambda^k (\frac{1}{2} \mathbf{v}^2 - \mathbf{v})$ ), and  $\lambda$  is the age of the SPGMN to control the learning pace,  $k$  means  $k$ -th iteration, while  $\mathbf{v} = [v_1, v_2, \dots, v_n]$  selects which samples involved into training data.

**Abnormal Student Detection Based on SPGMN.** Here, we introduce our ASD method based on SPGMN algorithm, i.e., SPGMN-ASD. We first train a SPGMN model for SLPP task on the training data. After enough iterations, we obtain a trained SPGMN model and the training loss of each student. Obviously, the set of losses for abnormal students (outliers) has clearly difference with the set of losses for normal students (majority of training samples). Based on this consideration, we first sort  $\alpha$  values (the loss of target student divide the sum of all losses). Then we randomly set one break point from the list of non-zero  $\alpha$  values and obtain two series of  $\alpha$ . Finally, if these two series cannot satisfy  $t$ -test, the corresponding students of series with larger  $\alpha$  values are detected as abnormal students.

### 3 Experiments

Here, we design our experiments based on real educational dataset from X-university, GPA-data. GPA-data consists of the registration information of 600 students, exam scores for courses they had enrolled (Each student totally enrolls about 70 to 80 courses.) and the background information of all courses. Here, data is constructed in the form of graphs (one graph represents a student). The nodes are courses that students had ever enrolled, and top-5 similar nodes are connected with non-weighted edges. Specially, 600 students are classified into 4 categories by GPAs, and  $GPA \in [1, 2, 3, 4]$ .

**Student Learning Performance Prediction.** To compare the performance of SPGMN and some compared baseline/SLPP methods, i.e., GMN [7], convolutional neural network (CNN) [4], K-Nearest Neighbor (KNN) [3], decision trees (DTs) [9], naïve Bayes (NBs) [10], we follow the experimental protocol in [11] and perform 10-fold cross-validation and report the mean accuracy over all folds.

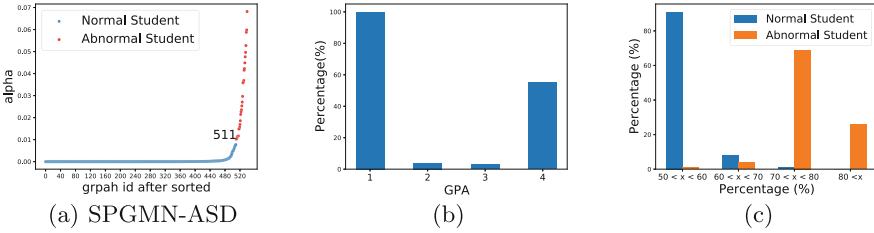
**Table 1.** SLPP results. Note, GPA- $k$  means the data of first  $k$  terms of GPA-data.

Methods	Datasets						
	GPA-1	GPA-2	GPA-3	GPA-4	GPA-5	GPA-6	GPA-7
KNN	44.83	42.83	40.17	45.00	40.67	36.83	37.50
DTs	49.83	50.50	59.50	59.83	60.67	61.00	60.33
NBs	35.17	40.83	55.83	54.17	51.00	50.50	50.33
CNN	69.83	68.83	70.00	72.33	74.17	77.83	79.67
GMN	76.00	79.50	89.17	91.17	92.67	93.84	94.50
SPGMN	<b>77.00</b>	<b>81.00</b>	<b>90.17</b>	<b>93.20</b>	<b>94.33</b>	<b>96.00</b>	<b>97.50</b>

We predict the final GPAs based on the data of the first  $k$  term ( $k \in [1, 7]$ ), i.e., GPA- $k$ . Table 1 shows that: (1) in all semesters, our SPGMN model consistently achieves better results w.r.t compared methods; (2) as semesters go, the advantage of SPGMN is amplified, i.e., SPGMN has strong robustness.

**Abnormal Students Detection.** Here, we test our ASD method on the training dataset of GPA-7. Specially, we randomly repeat SPGMN-ASD 400 times and consider the  $k$  student with highest frequency appeared as abnormal students. Note,  $k$  is the average number of results of 400 experiments.

Figure 1(a) suggests that: 1) the abnormal series and the normal series has a noticeable gap at 511-th point; 2) these two series come from two different distribution. Then, Fig. 1(b) indicates that GPA is not enough to evaluating students: 1) 100% of promising students were abnormal, i.e., they need more attention from teachers; 2) For the students whose GPA is 4, 56% of them



**Fig. 1.** (a): Results of ASD. (b): y-axis is the percent of all students who abnormal. (c): x is the percentage of major-related courses cut of all courses. y-axis is the percentage of student with different learning pattern cut of all students. (Color figure online)

were abnormal, i.e., those gifted students need more challenged works while the remaining 44% may not need.

By studying the 30 outliers, we found that they are not errors. Instead, it represents a noticeable event: the strategy of course selection (i.e., learning pattern) of abnormal students is different from normal students<sup>1</sup>. As shown in Fig. 1(c): 1) normal students (blue bars) are more likely to select non-major-related courses, e.g., *Basic of Finance* for students of materials science and technology; 2) abnormal students (orange bars) are more likely to choose major-related courses, e.g., *Database* for students of computer science; 3) when students enrolled more major-related courses, their learning patterns are more likely abnormal, i.e., learning patterns can be considered as a supplement of GPAs.

## 4 Conclusion

This paper combines SLPP task and ASD task for strong interpretability of ASD task. And we combine GMNs model and SPL for enhancing the robustness of GMNs for SLPP task and improving the precision of ASD task. Then, experiments verify the mentioned advancements. Finally, we propose the learning pattern as an supplement of GPA by analyzing the results of ASD experiment.

**Acknowledgement.** This research is funded by the National Natural Science Foundation of China (Grants No. 61802313, U1811262, 61772426), the Fundamental Research Funds for Central Universities (Grant No. G2018KY0301) and the education and teaching reform research project of Northwestern Polytechnical University (Grant No. 2020JGY23).

## References

1. Kuh, G.D., Kinzie, J., Buckley, J.A., Bridges, B.K., Hayek, J.C.: Piecing Together the Student Success Puzzle: Research, Propositions, and Recommendations: ASHE Higher Education Report, vol. 116. Wiley, Hoboken (2011)

<sup>1</sup> Here, we do not consider these required courses which students must enrolled.

2. Zhang, Y., Dai, H., Yun, Y., Liu, S., Lan, A., Shang, X.: Meta-knowledge dictionary learning on 1-bit response data for student knowledge diagnosis. *Knowl. Based Syst.* **205**, 106290 (2020)
3. Al-Shehri, H., et al.: Student performance prediction using support vector machine and k-nearest neighbor. In 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–4. IEEE (2017)
4. Ma, Y., Zong, J., Cui, C., Zhang, C., Yang, Q., Yin, Y.: Dual path convolutional neural network for student performance prediction. In: Cheng, R., Mamoulis, N., Sun, Y., Huang, X. (eds.) WISE 2020. LNCS, vol. 11881, pp. 133–146. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-34223-4\\_9](https://doi.org/10.1007/978-3-030-34223-4_9)
5. Zhang, Y., Yun, Y., Dai, H., Cui, J., Shang, X.: Graphs regularized robust matrix factorization and its application on student grade prediction. *Appl. Sci.* **10**(5), 1755 (2020)
6. Zhang, Y., An, R., Cui, J., Shang, X.: Undergraduate grade prediction in Chinese higher education using convolutional neural networks. In: 11th International Learning Analytics and Knowledge Conference, LAK 2021, pp. 462–468 (2021)
7. Khasahmadi, A.H., Hassani, K., Moradi, P., Lee, L., Morris, Q.: Memory-based graph networks (2020). arXiv preprint [arXiv:2002.09518](https://arxiv.org/abs/2002.09518)
8. Meng, D., Zhao, Q., Jiang, L.: A theoretical understanding of self-paced learning. *Inf. Sci.* **414**, 319–328 (2017)
9. Kabakchieva, D.: Student performance prediction by using data mining classification algorithms. *Int. J. Comput. Sci. Manag. Res.* **1**(4), 686–690 (2012)
10. Amra, I.A.A., Maghari, A.Y.: Students performance prediction using KNN and Naïve Bayesian. In: 2017 8th International Conference on Information Technology (ICIT), pp. 909–913. IEEE (2017)
11. Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. In: Advances in Neural Information Processing Systems, pp. 4800–4810 (2018)