NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

DEPARTMENT OF BUSINESS INFORMATICS

PROJECT

PREDICTING STUDENTS' GRADE USING GRAPH NEURAL NETWORKS

Iurii Gorbunov

Yokutkhon Naimkonova

Liliia Mohammad

Advisor:

Karpov Ilia,

Senior Lecturer of the Faculty of Computer Sciences

Moscow

2023

Abstract

The main idea of work is to improve educational progress at both institutional and individual levels. The knowledge tracing usage as a tool for advancing educational research and practice is promising. Predicting student grades and performance on exams could help the universities and students with insights to make changes towards a more productive and successful educational process. In addition, our proposal gives information on the importance of understanding which data most impacts student performance. Moreover, it is highlighted for universities to improve the quality of education and for students to pay attention to specific aspects of their studies.

Introduction

Knowledge tracking is the sphere where machine and deep learning models chase and evaluate students' educational progress. The main goal of this sphere is to implement algorithms for modeling students' knowledge and to improve educational progress not only at the level of the institution but also of the individual, facilitating more rapid changes in the educational process (Piech, C, etc., 2015). Thus, knowledge tracing represents a promising tool for advancing educational research and practice.

On the one hand, universities and educational platforms are interested in ensuring that students' academic performance is at a high level. On the other hand, the motivation of a student to have good grades is much more obvious, but also remains relevant. Predicting that the student's grades or their performance on the exam will be low, universities can offer him help, and the student himself can change his attitude to the educational process towards a more productive, and, ultimately, more successful one.

Moreover, it is worth considering how machine learning models work: based on a variety of different data, the final result is predicted, but not all data are equally influential on the result. In other words, which data have the greatest impact on student performance is also an important question, the answer to which will help universities improve the quality of education, and students pay attention to certain aspects of their studies.

Literature Review

The first attempts to model knowledge tracing process used Naïve Bayes algorithms (Corbett, A. T., & Anderson, J. R., 1994), decision trees and support vector machines to deal with classification problem. However, as time goes on and computing power improves – the better results in knowledge tracing were proposed by deep learning algorithms (Piech, C, etc., 2015), more specifically, by neural networks (Mengash, H.A., 2020). Unfortunately, presented universal models do not take into consideration individual learning features of individuals.

The study of College of Electronic and Information Engineering, Tongji University, Shanghai on topic “Graph Neural Network for Senior High Student’s Grade Prediction” used GNNs to predict students' academic performance in online courses. The study used data from 10,765 students who completed 36 online courses on the edX platform. The data included various demographic and academic variables, such as students' age, gender, and previous academic achievements. The review used a dataset containing data on understudies' socioeconomics, earlier course execution, and informal community cooperation inside a specific course.

The specialists originally analyzed the interpersonal organization information and found that understudies who had more associations with their friends would in general perform better scholastically. They then, at that point, prepared a GNN on this dataset, utilizing it to foresee understudies' last course grades.The results showed that the GNN had the option to precisely anticipate understudies' grades with a typical mistake pace of just 2.63 brings up of 100. The specialists likewise observed that the GNN had the option to recognize which highlights of the dataset were generally significant for anticipating execution, which included interpersonal organization centrality, earlier course grades, and orientation. (Yu, 2022)

Generally, the review features under topic “Study-GNN: A Novel Pipeline for Student Performance”. Prediction Based on Multi-Topology Graph Neural Networks the potential for GNNs to be used as a device for anticipating understudies' scholarly performance, especially while thinking about interpersonal organization collaborations as an element. This could have significant ramifications for instructive foundations hoping to recognize understudies who might be in danger of falling behind and give designated mediations to work on their presentation.

The pipeline takes into account the complex relationships between different educational entities, such as students, courses, and instructors, and builds a graph-based representation of this information. The multi-topology aspect of the approach allows for the incorporation of different graph topologies to better capture these relationships. (Ming Li, 2022)

The model is trained using a combination of supervised and unsupervised learning techniques, which enables it to learn from both labeled and unlabeled data. This approach allows for improved accuracy in predicting student performance across a variety of educational settings.

These studies demonstrate the potential of GNNs for predicting students' grades based on their academic and social interactions. They show that by modeling the relationships between students and courses as a graph, GNNs can capture complex patterns of interactions and provide accurate predictions of academic performance. However, further research is needed to explore how GNNs can be used to support personalized learning and early interventions for students at risk of falling behind.

Finally, the article “Online Academic Course Performance Prediction usingRelational Graph Convolutional Neural Network” by authors Hamid Karimi, Tyler Derr, Jiangtao Huang, Jiliang Tang gives analysis of using the GNN to Deep Online Performance Evaluation (DOPE), which firstly models the student course relations in an online system as a knowledge graph, then utilizes an advanced graph neural network to extract course and student embeddings, harnesses a recurrent neural network to encode the system's temporal student behavioral data, and ultimately predicts a student's performance in a given course. Comprehensive experiments on six online courses verify the effectiveness of DOPE across multiple settings against representative baseline methods. Furthermore, we perform ablation feature analysis on the student behavioral features to better understand the inner workings of DOPE. (Hamid Karimi, 2022)

Main part

The structure of students’ course progress data can be represented similarly to the graph structure: G = (V, E, A). The necessary knowledge for the successful mastering of the program are decomposed into N educational concepts, known as vertices $V = \{v_1, \cdot \cdot \cdot, v_N\}$, and these concepts have dependency relationships – edges $E \subseteq V \times V$. The degree of dependency between edges is defined by the adjacency matrix $A \in R_N \times N$ (Nakagawa, H. etc., 2019).

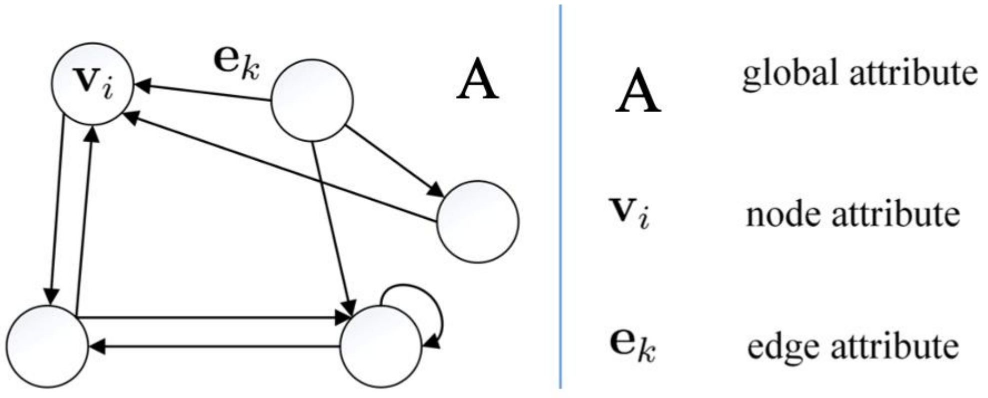

Figure 1. Graph network block

One of the approaches of implementing graph structure in knowledge tracking is building graph neural networks (GNN) model. It was widely used in different spheres: from classification to prediction and recommendation systems (graph attention multi-layer). By storing layers, GNNs are able to learn node representations with the help of utilizing information from the node's neighborhoods in the layer (Yu, Y., etc., 2022).

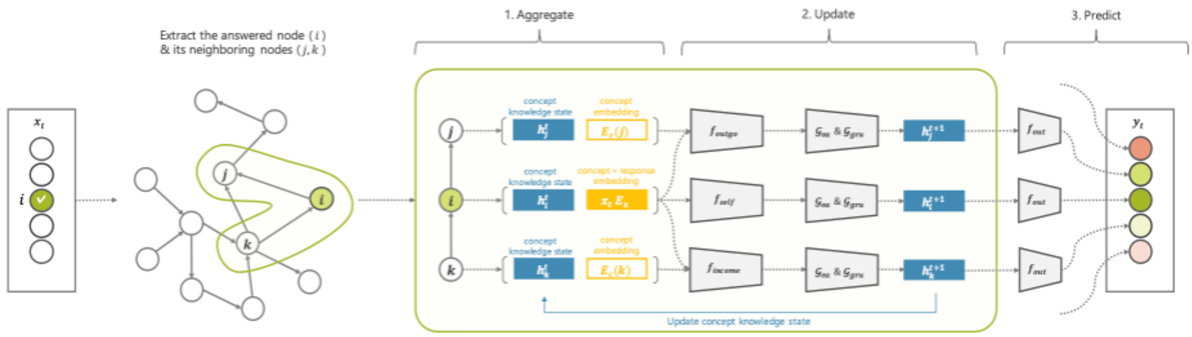

Figure 2. The architecture of graph knowledge tracking with multi-layer perceptron.

The process of implementing neural network to graph structured data is presented in figure 2. The first step of the model – to *aggregate* knowledge states and any additional embeddings for knowledge concept *i* and concepts $j \in N_i$ in the neighborhood. The second one is to *update* knowledge states based on aggregated features and the structure of the graph, where f_{self} states for multi-layer perceptron (MLP), G_{ea} states for erase-add gate, and G_{gru} – for gated recurrent unit gate. Finally, the model estimates the probability of a student answering concepts correctly at the next time steps (Nakagawa, H. etc., 2019).

Anticipated SNA Methods

Expected Results

Conclusion (optional)

References

1. Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4, 253-278.
2. Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In IEEE/WIC/ACM International Conference on Web Intelligence (pp. 156-163).

3. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. Advances in neural information processing systems, 28.
4. Yu, Y., Fan, J., Xian, Y., & Wang, Z. (2022). Graph Neural Network for Senior High Student’s Grade Prediction. Applied Sciences, 12(8), 3881.
5. Hamid Karimi, T. D. (2022). Online Academic Course Performance Prediction using Relational Graph Convolutional Neural Network. Nanning Normal University, China.
6. Ming Li, X. W. (2022). Study-GNN: A Novel Pipeline for Student Performance Prediction Based on Multi-Topology Graph Neural Networks. Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua 321004, China;. Получено из Study-GNN: A.
7. Yu, Y. F. (2022). Graph Neural Network for Senior High Student’s Grade Prediction. Applied Sciences.

Appendices (optional)