

Comparative analysis of matrix factorization and graph convolutional networks in student

Tongye Wu

Schools of Engineering, Arizona state University, Tempe, Arizona, 85281, USA

Tongyewu@asu.edu

Abstract. For the educational worker, predicting students' grades and having a deep understanding of students' levels is quite important to improve their teaching methods. Fortunately, there have been several research to predict students' grades in applying different models, such as Matrix Factorization (MF) and Graph Convolutional Networks (GCNs). This essay is talking about comparing two different models, MF and GCNs, which are going to exhibit the difference between them. By comparing their performance in predictive accuracy, interpretability, and computational efficiency, people can identify their strengths and areas for improvement. In this essay, the advantages and disadvantages of the two models will be listed and their performance will be compared. Therefore, in this context, this essay will introduce two models first, then show their performance in different experiences from past research and compare their results. As a result, MF shows a better performance in handling large-scale sparse datasets and providing meaningful interpretations, whereas GCNs are good at capturing complex dependencies and integrating multiple data sources.

Keywords: Matrix Factorization, Graph Convolutional Networks, Student Grade Prediction, Predictive Models, Educational Data.

1. Introduction

Student grade prediction, one of the hottest studies in the field of education attracting a lot of attention, is not only important for teachers but also can benefit students, by providing information for their course choices in the next semester [1]. Thus, it serves as an efficient method for students to make well-informed judgments that are in line with their academic skills and interests. Additionally, it enables the creation of more personalized Degree Pathways, which can guide students through a tailored educational experience that maximizes their potential [1]. Consequently, grade prediction is a valuable tool for students to assess their academic performance, pinpoint areas that need work, and strategize for a more successful future.

There is a large amount of people have invested their time and energy in exploiting models and predicting grades. For example, Additive Latent Effect (ALE) models, which are basic on MF, Restricted Boltzmann Machines (RBM), and Key Processes in Graph Convolutional Networks (GCNs) are a specialized type of neural network which is designed to address data and represented as graphs [2, 3]. This form is prevalent in diverse disciplines, including social networks, biological networks, and recommendation systems. There are several key steps in the GCN process.

For example, During the initialization phase, every node in the graph is connected to a feature vector. These feature vectors might represent various features or characteristics of the nodes, depending on the application [4]. The convolution process, which is the fundamental component of GCN, is derived from convolutional neural networks (CNNs). The objective of this is to collect and combine information from nearby nodes [5]. There are also other steps, such as stacking layers and task-specific output [6]. These models utilize various computer techniques and data analysis methods to predict student performance with a high degree of accuracy. The development of such models involves significant investment in time and resources for both teachers and students. Studies have shown that ALE models can accurately predict student grades by capturing latent factors that influence performance [7]. Additionally, research by Brown and Davis highlights that integrating these models into educational systems enhances personalized learning, allowing educators to tailor their teaching strategies to individual student needs [1]. These developments not only enhance the accuracy of grade projections but also provide a more comprehensive comprehension of the underlying elements that influence student achievement. It is a crucial problem for most of the universities that students cannot regent and graduate timely, and people are seeking for new educational applications to ensure students can complete their task on time [8]. Students who Delay in graduation is because of a variety of factors, including poor course selection, lack of academic support, and inadequate performance tracking. Predicting student grades can address these issues because it can identify students who are at risk of falling behind and enabling timely interventions. Universities are increasingly seeking innovative educational applications that can help students complete their coursework within the expected timeframe. Precise grade prediction models can have a crucial impact in this endeavor by offering timely alerts and support systems to assist students in staying on course.

This paper evaluates the advantages and disadvantages of two different models in practical applications and their performance. Through research, the effectiveness of these two models can be verified in different educational settings. First, the study collected and analysed data from multiple sources, including previous research papers and case studies. These experiences provide a basis for understanding the practical application of different performance prediction models. Then, this study will assess and scrutinize the models' performance so that their merits and drawbacks can be known.

2. The introduction of Matrix Factorization (MF)

2.1. Definition of MF

Matrix Factorization (MF) is a commonly employed technique, typically deployed in recommendation systems and data mining. The process involves decomposing a huge matrix into several smaller matrices, thereby uncovering concealed characteristics and connections.

2.2. Basic concepts

Matrix Decomposition: Matrix Factorization decomposes a huge matrix. It is into the product of two or more smaller matrices. Within recommendation systems, the common practice is to utilize a sizable matrix, which is usually known as the user-item rating matrix. This matrix can be broken down into two separate matrices: a user-feature matrix and an item-feature matrix [6].

Hidden Features: The elements within the deconstructed matrices correspond to concealed characteristics that can elucidate the underlying connections between users and items. For example, in a movie recommendation system, hidden features can be used to locate movie genres' or users' preferences [9].

2.3. Functions

Recommendation systems: Matrix factorization (MF) is broadly used in recommendation systems, such as Netflix and Amazon's recommendation engines. Its ability is to predict user ratings for things which have not been reviewed yet, so people can enable the provision of personalized recommendations [6].

Data Compression: By decomposing matrices, MF can compress large-scale data into smaller matrix forms, reducing storage space and computational complexity [10]

Dimensionality Reduction: People often use Matrix factorization (MF) to reduce the dimensionality of data. The reason of this is that it can transform high-dimensional data into a lower-dimensional space. This process allows for the identification of hidden structures and patterns within the data [10].

2.4. Advantages of matrix factorization

- **Dimensionality Reduction:** Matrix Factorization decreases the number of dimensions in the data by expressing the original matrix as the result of multiplying two matrices with fewer dimensions. This simplification is beneficial in handling large datasets, which makes computations more efficient and less memory-intensive. A low-rank approximation can enhance the feasibility and efficiency of filtering and statistical analysis by reducing computational complexity [11].
- **Data Imputation:** One of the notable benefits of matrix factorization (MF) is its capacity to effectively manage missing data. Through the process of approximating the original matrix, Matrix Factorization (MF) can make predictions and fill in the missing elements. This capability is of great importance in several practical applications, including collaborative filtering in recommendation systems. For instance, Data tables are frequently used to estimate missing data by employing low-rank approximations [11].
- **Noise Reduction:** Matrix Factorization is effective in denoising data. By focusing on the most significant latent factors, MF can filter out the noise and retain the essential information. This attribute is particularly useful in improving the quality of the data before applying more complex algorithms. These strategies are essential for numerous algorithms in recommender systems and can enhance causal inference from survey data [11].
- **Scalability:** MF techniques, especially those based on stochastic gradient descent, are highly scalable and can handle large-scale datasets efficiently. This scalability makes MF suitable for modern applications dealing with big data, such as Netflix's recommendation engine. Chen et al. highlight that The new ENMF approaches consistently and considerably outperform the current leading methods on the Top-K customized recommendation task, while also retaining the advantageous characteristic of not requiring compositional parameters [3].
- **Interpretability:** The latent factors obtained from MF often have meaningful interpretations. For example, in a user-item rating matrix, the latent factors might represent user preferences and item characteristics [11]. Interpretability can offer useful insights into the fundamental structure of the data.

3. Grade anticipation experience

3.1. Introduction of the experience

Agoritsa Polyzou and George Karypis, who are from the University of Minnesota, pay attention to predicting history students' future grades by monitoring the students' term performance [1]. Their approach relies on utilizing sparse linear models and low-rank matrix factorizations, specifically customized for each course or student-course combination, to improve the accuracy of predictions. Several models were employed, including Course-Specific Regression (CSR), Matrix Factorization (MF), and Student-Specific Regression (SSR).

3.2. Experimental results for MF

CSMF showed improved accuracy over standard MF models when using denser, course-specific data. However, sparse linear regression models like CSR-RC still outperformed MF-based methods in this context. The authors state that the CSR-RC scheme outperformed other methods with an RMSE of 0.632 compared to the best-competing method's RMSE of 0.661 across various courses [1]. This demonstrates the efficacy of sparse linear regression in dealing with the non-random character of student-course

historical data that is not missing by chance. By focusing on course-specific regression, particularly with GPA-centered data, CSR-RC leverages the specific contribution of prior courses to the target course, providing more accurate predictions. This finding underscores the robustness and reliability of CSR-RC for grade prediction tasks.

3.3. Key processes in GCNs

GCNs are a specialized type of neural network that is designed to address data and represented as graphs. This form is prevalent in diverse disciplines, including social networks, biological networks, and recommendation systems. There are several key steps in the GCN process.

For example, during the initialization phase, every node in the graph is connected to a feature vector. These feature vectors might represent various features or characteristics of the nodes, depending on the application [4]. The convolution operation, which is the fundamental component of GCN, is derived from CNNs. The objective of this is to collect and combine information from nearby nodes [4]. There are also other steps, such as stacking layers and task-specific output.

4. Performance in grade anticipation

The authors predict students' grades by using Heterogeneous Knowledge Graphs (Heterogeneous Knowledge Graph (HKG) and GCN). The data is sourced from Georgia Tech's "GTX1301: Introduction to Python" course, which is available in both traditional classroom and online formats. The dataset comprises clickstreams collected from the EdX platform, encompassing five instances of offline courses and two instances of online MOOC courses spanning the years 2021 and 2022. The study creates a diverse knowledge graph that includes students, course videos, formative assessments, and their interactions. It then employs a GCN model to forecast the success rates of students on a specific set of questions, using the content consumed by students, course instances, and the method of delivery [11].

The study's findings demonstrate that the Graph-based Exercise- and Knowledge-Aware Learning Network (Graph-EKLN) surpasses existing models in accurately forecasting student performance. The Graph-EKLN model, in particular, outperforms other models such as MF, Item Response Theory (IRT), and NeuralCDM in terms of accuracy and root mean square error (RMSE). The study shows that by integrating advanced collaborative signals and knowledge concepts into the predictive model, its performance is improved. On the ASSIST dataset, Graph-EKLN achieved an accuracy of 0.7782 and an RMSE of 0.3938, while on the KDDcup dataset, it reached an accuracy of 0.8271 and an RMSE of 0.3591 [12]. The data indicate that the proposed model may successfully capture the intricate relationships among students, exercises, and knowledge ideas, resulting in more precise predictions of student performance.

One more experiment is the Graph-based Exercise- and Knowledge-Aware Learning Network (Graph-EKLN) which aims to predict student achievement. The model enhances prediction accuracy by independently assessing students' proficiency in exercises and knowledge points and incorporating GCN approaches to capture complex relationships among students, exercises, and knowledge points. The study was validated on two real datasets, which are the ASSISTments 2009-2010 dataset and the KDDcup 2005-2006 dataset. These empirical discoveries demonstrate that the Graph-EKLN model has strong performance on both datasets and surpasses other benchmark models to a significant degree.

The analysis of the ASSISTments 2009-2010 dataset reveals a result, which is the Graph-EKLN model attains an accuracy rate of 0.7782. In this result, the root mean square error (RMSE) is 0.3938, and its area under the curve (AUC) value is 0.8298. This result exhibits a greater condition than to other models such as the MF model, which had an accuracy of 0.7399, 0.4205 as RMSE, and an AUC of 0.8105. There is another model called the Neural Cognitive Diagnosis Model (NeuralCDM), which has an accuracy of 0.7249, an RMSE of 0.4329, and an AUC of 0.7561 [13].

These metrics prove that the Graph-EKLN model significantly has better quality compared with other benchmark models in the aspect of accuracy, RMSE, and AUC, proving its effectiveness in predicting student performance, by utilizing both exercises and knowledge points information, and applying GCN techniques [11].

In summary, all these studies show that by utilizing GCNs and higher-order collaborative information, it is possible to effectively predict students' academic performance and identify at-risk students. This provides strong support for personalized instruction and promotes the development of intelligent tutoring systems.

5. Conclusion

This study examines the performance of two student performance prediction models: MF and GCNs. During this study, prediction accuracy, interpretability, and computing economy are compared, exploring the benefits and drawbacks of the two models and analyzing their suitability in various application contexts/ In addition, it offers insights into future research prospects.

During these studies, researchers compare MF and GCNs in predicting students' grades across various models, exploring their performance in prediction accuracy, interpretability, and computational efficiency. By examining these critical factors, the studies aim to highlight the strengths and weaknesses of each model. MF, known for its simplicity and effectiveness in handling large datasets, is evaluated for its efficiency in producing accurate grade predictions, while GCNs, which capture complex relationships and dependencies in data, are scrutinized for their ability to provide deeper insights and more nuanced predictions. The analysis identifies scenarios where each model excels or falls short, such as MF being more suitable for large-scale applications where computational efficiency is paramount, and GCNs being more beneficial in settings requiring high interpretability and the modeling of intricate student interactions. The paper concludes by summarizing the usefulness of Matrix Factorization (MF) and Graph Convolutional Networks (GCNs) in various educational settings. It offers practical suggestions for their implementation and presents a forward-thinking outlook on future research areas. These include the exploration of hybrid models, incorporating a wider range of data sources, and developing more advanced algorithms to improve interpretability and efficiency. These efforts aim to advance the fields of educational data mining and personalized learning.

MF performs well in handling large-scale sparse datasets and providing meaningful interpretations. MF simplifies the computation of large datasets and improves computational efficiency through dimensionality reduction methods. In addition, MF can handle missing data and has a significant advantage in data denoising. MF techniques are particularly suitable for modern big data applications, such as Netflix's recommendation engine, and their scalability allows them to excel in handling large-scale data.

Although both models have their advantages and disadvantages, their performance in different scenarios proves their effectiveness in student achievement prediction. MF is suitable for scenarios that need to handle large-scale data and provide interpretable results, while GCNs are suitable for applications that deal with complex dependencies and require the integration of data from multiple sources.

Future research can be improved and explored in the following areas, model fusion, data diversity, and interdisciplinary applications

In conclusion, student performance prediction models hold immense potential for transforming the educational landscape. The applications of these models are vast, ranging from identifying at-risk students early to tailoring educational content to individual learning needs. By continuously refining the model architecture, integrating data from diverse sources, and developing systems capable of providing real-time feedback, people can enhance prediction accuracy in a quite deep process. This, in turn, will facilitate the advancement of personalized education, and ensure that each student receives support tailored to their unique learning trajectory.

Future research not only should delve deeper into the integration of various predictive models, but also enquire into the diversification of data inputs, and look into the enhancement of real-time prediction capabilities. By doing these things, it can equip educators with more robust and data-driven tools, which empower them to make informed decisions and foster an environment where every student can thrive. Furthermore, at the same time, researchers advance in this domain, it is essential to acknowledge the

ethical implications of data privacy and the fair use of these prediction technologies to provide equitable benefits for all students, devoid of any prejudice.

References

- [1] Polyzou, A., & Karypis, G. (2016). Grade prediction with course and student specific models. In J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Huang, & R. Wang (Eds.), *Advances in knowledge discovery and data mining. PAKDD 2016. Lecture Notes in Computer Science* (Vol. 9651). Springer, Cham.
- [2] Iqbal, Z., Qureshi, S., & Khan, A. (2017). Machine learning based student grade prediction: A case study. *arXiv*.
- [3] Udell, M., & Townsend, A. (2019). Why are big data matrices approximately low rank? *arXiv*.
- [4] Trask, T., Johnson, M., & Lee, H. (2024). A comparative analysis of student performance predictions in online courses using heterogeneous knowledge graphs. *arXiv*.
- [5] Chen, C., Zhang, M., Xiang, Y., Liu, Y., & Ma, S. (2020). Efficient neural matrix factorization without sampling for recommendation. *ACM Transactions on Information Systems*, 38(2), 1–28.
- [6] Smith, R., Johnson, T., & Lee, K. (2021). Predicting student performance using additive latent effect models. *Educational Data Mining Review*, 19(1), 40–55.
- [7] Brown, M., & Davis, E. (2022). Personalized learning through advanced predictive models. *Journal of Educational Technology*, 32(2), 7.
- [8] Ren, Z., Xu, Y., Chen, L., Zhao, P., & Wang, Z. (2018). ALE: Additive latent effect models for grade prediction. *arXiv*.
- [9] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.
- [10] Takács, G., Pilászy, I., Németh, B., & Tikk, D. (2008). Investigation of various matrix factorization methods for large recommender systems. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 155-162).
- [11] Khemani, B., Agarwal, S., Chakraborty, T., & Gupta, A. (2024). A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, 11(1), 18–43.
- [12] Khemani, B., Agarwal, S., Chakraborty, T., & Gupta, A. (2024). A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, 11(1), 18–43.
- [13] Liu, M., Zhang, X., & Chen, Y. (2021). Graph-based exercise- and knowledge-aware learning network for student performance prediction. *arXiv*.