Learning Analytics based on Multilayer Behavior Fusion

Yu Yang, Jiannong Cao, Jiaxing Shen, Ruosong Yang, Zhiyuan Wen
Department of Computing,
The Hong Kong Polytechnic University, HK SAR
{csjcao,jiaxshen, csyyang, csryang, cszwen}@comp.polyu.edu.hk

Abstract. Learning analytics is the measurement, collection, and analysis of data about learners and their contexts for the purposes of understanding and optimizing the process of learning and the underlying environment. Due to the complex nature of the learning process, existing works mostly focus on the modeling and analysis of single learning behavior and thus bears limited capacity in achieving good performance and interpretability of predictive tasks. We propose a research framework for learning analytics based on multilayer behavior fusion which achieves significantly better performance in various tasks including at-risk student prediction. Results of extensive evaluation on thousands of students demonstrate the effectiveness of multilayer behavior fusion. We will report the insights about mining learning behaviors at different layers including physical, social and mental layers from the data collected from multiple sources. We will also describe the quantitative relationships between these behaviors and the students' learning performance.

Keywords: Learning analytics · Multilayer behavior extraction · At-risk student prediction · Automatic text scoring

1 Introduction

Proper education is the foundation of civilization, happiness, and success. Education has thus been an enduring and significant topic at different times. The past decades have witnessed the rapid advancement of wireless sensing, big data, and artificial intelligence. Novel pedagogical improvements have been achieved significantly via community-based learning environments where learners could learn in online communities like discussion forums and various Learning Management Systems (LMS) like Blackboard. These emerging learning paradigms are the foundation of Learning Analytics. The huge amount of data generated through online and offline activities make it possible to trace the learning processes and analyze their relationships with learning outcomes quantitatively. Specifically, learning analytics is defined as "the measurement, collection, and analysis of data about learners and their contexts for the purposes of understanding and optimizing the process of learning and the underlying environment" [1].

Although applications in learning analytics share a similar purpose which is to tailor educational opportunities to the individual learner's need, they are quite diverse ranging from performance prediction to course recommendation [2, 3]. For instance, Purdue University exploited data from LMS of a certain course to predict which students may be struggling academically and to provide proactive intervention. The assumption is the students' effort measured by participation in LMS could partially explain academic success. Another example is The University of Alabama improved student retention using a predictive model for at-risk students based on the large data set of their demographics [4].

In various learning analytics applications, different data modalities including text, video, and spatiotemporal data are used to model learners' learning behavior. However, most of them focus on the behavior in a single layer, which bears a limited capacity to model the underlying learning processes that are complex and dynamic in nature. For example, the LMS data reveal the underlying patterns of how students participate in the course. The way students use the LMS like how long (timespan) and how often (frequency) they use certain functionality is closely correlated with their learning performance. The statistical features of timespan and frequency belong to physical behavior. However, learners with the same physical behavior do not necessarily get similar outcomes due to the dynamics in human behavior [21]. Every single human activity takes place within a context. Learning is not an exception. Both relational context and mental context play vital roles in understanding human behavior. The relational context captures social influence and social interaction between learners while the mental context includes individuals' emotion, perception, and motivation.

In this paper, we extract behavior in multiple layers including physical, social, and mental layers which could model the underlying learning process in a more comprehensive way. The vision, however, entails two grand challenges when applied to reality. The first challenge is how to design multilayer behavior features? Most of the related works focus on physical behavior features. How to model the social interaction and mental status of learners remain open challenges. The second challenge is how to fuse the behavior feature from multiple layers? Given features from different behavior, they are usually of different modalities and different granularity which make the fusion task difficult and challenging.

We propose a general research framework indicating how multilayer features are extracted and fused for two learning analytics applications. For the extraction challenge, instead of directly measuring the social interaction in the physical world that is barely possible, we use alternative ways to approximate the social network. For example, both co-occurrences in the physical world and quotation in the cyber world are indicators of social interaction [5]. We also resort to the regularity of different behavior as the measurement of personal characteristics [6]. For the fusion challenge, we propose two types of fusion including feature level fusion and model level fusion as illustrated in two different learning applications. The effectiveness is evaluated in both applications. The results indicate the propose research framework could significantly improve application performance with the help of multilayer behavior fusion.

The remainder of the paper is organized as follows. Section 2 gives an overview of the proposed research framework. Section 3 and Section 4 are two case studies of learning applications which illustrate how multilayer behavior features are extracted and fused. The last section concludes the whole paper.

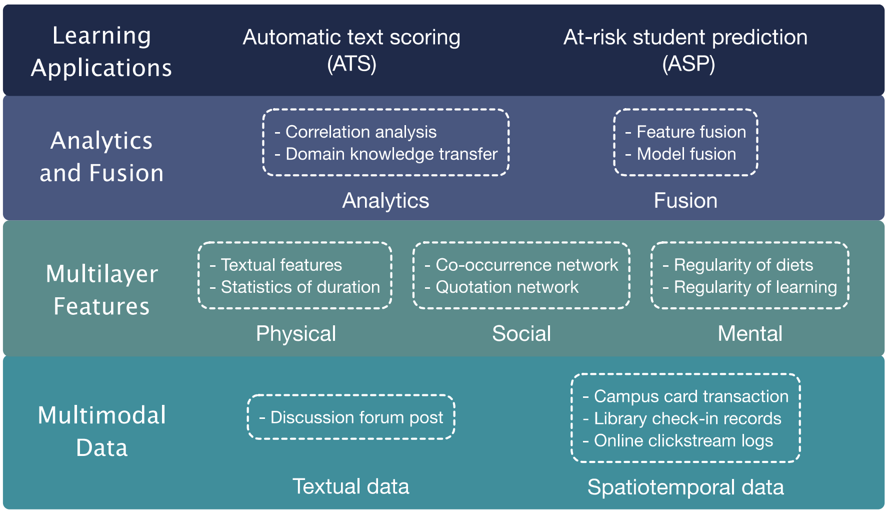

Fig. 1. The proposed research framework.

2 Overview

To comprehensively model the learning process, we propose the following research framework as shown in Figure 1. This framework illustrates how we process typical multimodal data for significant learning analytics applications of automatic text scoring and at-risk student prediction. Instead of extracting physical behavior only, we extract and fuse social and mental behavior.

The Multimodal Data Layer shows the textual data and spatiotemporal data we processed. Specifically, textual data refer to the online discussion forum posts from students and spatiotemporal data include transaction data of daily consumption using campus cards, check-in histories of the library, and operation logs in LMS. In the next layer, we managed to extract multilayer features from physical, social, and mental levels. Physical features are mostly conventional features like statistical features of timespan, frequency of certain activities and textual features from the forum posts. Social features include networks inferred from co-occurrence and quotation. Mental features are the regularity of different behavior. Then in Analytics and Fusion Layer, we transfer domain knowledge and fuse features from different levels for the top layer applications including automatic text scoring (ATS) and at-risk student prediction (ATP).

ATS is to automatically mark the score of given online forum posts. The main focus is textual data where we extract both physical and social features. We consider not only the writing quality but also the quotation relationships between posts using model level fusion. ATP is to predict students that will be academically at-risk. The main inputs are spatiotemporal data from which we extract physical, social, and mental features. The feature-level fusion of all the features is quite effective in predicting at-risk students owing to the capacity of capturing the dynamics of learning processes.

3 Case Study 1: Automated Online Forum Posts Scoring

3.1 Introduction to Automated Online Forum Posts Scoring

With the rapid development of the Internet, online courses ware is spreading worldwide at exponential speed. Numerous colleges and universities have offered fully online or hybrid courses combining online instruction with face-to-face teaching. In 2011, a study from the Pew Research Center reported that, in 2010-11 academic year, 89 percent of four-year colleges and universities offered courses taught fully online, or hybrid/blended online, or other forms of distance instruction. In 2013, 32 percent of all students would enroll in higher education took at least one online course [7]. Meanwhile, instructors start to assign homework online and ask students to submit their homework online. With more and more electrical homework, the requirement of automated grading becomes increasingly urgent. It is a heavy burden for each instructor to mark hundreds of homework within limited time. Besides, unlike objective questions that have explicit answers, the answers of subjective assignments only provide some guidelines so that multiple instructors may give different scores for the same assignment due to different judgments. Last but not least, it is difficult to avoid preferences of instructors' tastes so that some assignments may gain higher scores than other assignments marked by the same instructors. To address the aforementioned issues, it is essential to mark the assignments automatically.

For subjective assignments, textual answers are the most popular way to show the arguments, introduce the method and procedure. Therefore, Natural Language Processing is adopted to analyze those textual answers where researchers largely focus on two types of problems. The first one is Automated Essay Scoring [8] which mainly concentrates on single long text, such as the composition or academic essays. These texts have hundreds of words and were grades by the instructors with a score according to their writing quality. There are usually significantly concrete criteria as guidelines to guide the instructors' marking. The major challenge here is how to represent the writing quality of the long text. The other is Automated Short Answer Grading (ASAG) [9] which pays more attention to the correctness of the student answers. In this task, a question and a correct answer are usually given by the instructors, and the answers are very short, usually one or two sentences. So, the key step to solve this task is to match the consistency between the correct answer and the students' answers.

In online education, online forum is widely used by both educators and students, since asynchronous, threaded discussions can be effective in creating a collaborative learning environment [10]. It benefits online learners via reducing their dropout rates, increasing their performance and course satisfaction, as well as helping and learning from each other [11]. Thus, instructors, especially in the field of social science, usually assign some open topics for students to discuss online as homework. It is inevitable that the instructors are required to grade students' performance by reading all students' posts. However, unlike the aforementioned two tasks, the online discussions show quotation relationships, some posts are more likely to be quoted such that the contents in these posts may carries more significant information. Posts with more quotations could reveal students' innovations since many students are interested in and discuss them. To summarize, this task is to automatically mark all the posts of each student not only considering the writing quality of the posts themselves, but also the quotation relationship among these students.

3.2 Model Fusion of Physical Features and Social Features

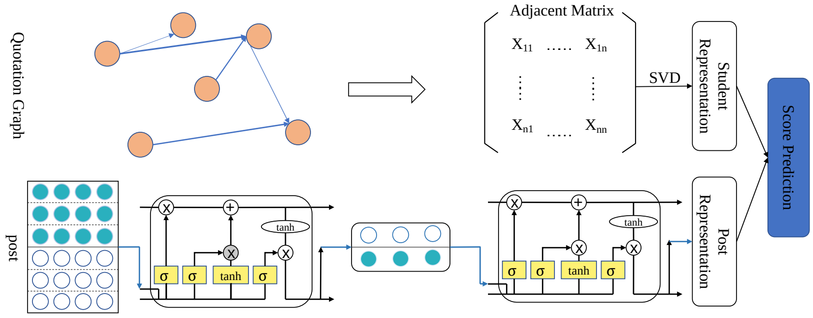

Fig. 2. Methodology framework of automated online forum posts scoring.

We proposed a new model as shown in Figure 2 to combine the measurement of writing quality and the topology of quotation relationship to grade the students' posts. To evaluating the writing quality, a hierarchical RNN model [12] is used to learn post representations that contain syntactic, semantic and coherence information. Besides, student representation is learned to capture the topology information from the quotation graph. More specifically, to learn the post representation, each post is separated into several sentences, and each sentence is separated into several words. Then, a Long Short-Term Memory (LSTM) network [13] is used to compose a sequence of words to learn sentence representations. Furthermore, another LSTM network is utilized to compose a sequence of sentences to learn post representations. As for student representation, refer to recent network embedding models, a quotation graph is constructed according to the quotation relationship between students. Then, an adjacent matrix is constructed according to the quotation graph. Finally, student representation could be learned via matrix factorization (Singular Value Decomposition) of the constructed adjacent matrix. In total, with learned post representation from text and student representation from the quotation graph, two features are combined to predict the score of the student's post.

3.3 Results of Model Fusion in Automated Online Forum Posts Scoring

We construct a dataset cooperated with Department of Applied Social Science in our university, and two types of evaluation metrics are used, namely correlation measurements including Quadratic Weight Kappa (QWK), Spearman Correlation Coefficient (SCC), and Pearson Correlation Coefficient (PCC), and residuals-based measurements, such as Rooted Mean Square Error (RMSE). As shown in **Table 1**, we show the experimental results of only using neural network (NN) and using both neural network and matrix factorization (NN + MF). Since the quotation relationship is very sparse and matrix factorization is too simple to learn the quotation relationship topology. The performance of utilizing quotation relationship is lower than only using neural network. However, With the extra quotation relationship, the model gains better QWK values, which shows the effectiveness of quotation relationship.

Table 1. Experiment Results.

Model	QWK	SCC	PCC	RMSE
NN	0.405	0.445	0.452	5.60
NN + FM	**0.417**	**0.430**	**0.439**	**5.96**

4 Case Study 2: Early Prediction of At-risk Students

4.1 Introduction to At-risk Student Early Prediction

Early predicting students at risk (STAR) is an effective and significant means of timely prevention of dropout and suicide. STAR are students who require temporary or ongoing intervention for achieving academic success [14]. Universities usually identify STAR by their academic performance which is sometimes too late to intervene. Existing works predict STAR from either online or offline learning behaviors [15,16,17]. However, neither of them is comprehensive enough to capture the whole learning processes and lead to unsatisfied prediction accuracy. For example, some students may prefer learning online but rarely attend face-to-face lectures. Thus, their offline learning behaviors are inactive which introduces biases in prediction if the whole learning process is not captured.

We aim to identify STAR before the end of a semester using multilayer behavior fusion from both online and offline learning activities. We define STAR as students whose Grade Point Average (GPA) is below 2.0 in a semester. The online learning behaviors are collected from the click-stream logs in the Blackboard, a learning management system (LMS), while the offline one comes from the library check-in records. There three major challenges to be tackled. (1) The number of STAR is far less than that of normal students such that STAR prediction is an extreme label-imbalance classification problem. (2) Comparing to the click-stream traces in LMS, the library check-in records are much sparser causing data density imbalance issues while data fusion. (3) STAR are usually inactive at the beginning of a semester so that their behavior traces are far less than enough for accurate early prediction.

To solve the aforementioned challenges, we propose EPARS for early predicting at-risk students. With the observation that study routines of good students are periodical [18] and STAR usually have more drop-out friends [19], EPARS extracts students' learning regularity patterns by a multi-scale bag-of-regularity approach and embeds their social homophily to accurately predict STAR. The experimental results show that EPARS outperforms baselines and achieves over 61% prediction accuracy in the first week of the semester.

4.2 Feature Fusion of Physical, Social, and Mental Behavior

To encode the regularity as features, we propose multi-scale bag-of-regularity to extract the repeated patterns of learning behaviors in multi-scale manners. First, we represent the learning behaviors as a binary sequence, then we sample subsequences at every nonzero element with length $\ell = 2 + (s - 1) \times z$ where integer s is the scale and z is the step-size between scales. These subsequences actually carry students' behavior patterns. Because regularity is the repeat of the behavior patterns, we filter out those only appearing once and count the number of occurrences of all possible behavior patterns in every scale as features. With this approach, which is robust for sparse data, the regularity features extracted from dense LSM data and sparse library check-ins are in the same scale-space so that it can well solve the challenge of data density imbalance. Figure 3 shows the average occurrence number of each library check-in pattern between STAR and normal students. The horizontal axis represents the library check-in patterns at scale 1 to 4. For example, pattern 110 represents a three-day pattern of students' library check-in behavior in which they continuously go to the library for first 2 days but not go there on the third day. The patterns at scale 1 exactly is the total number of library check-ins. This figure indicates that STAR have less continuous library studies than normal students.

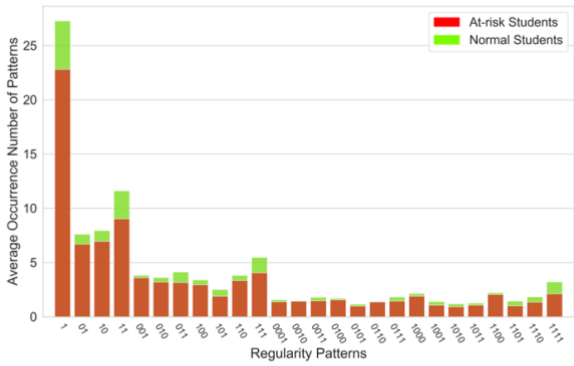

Fig. 3. Regularity patterns of at-risk students and normal students.

To supplement the lack of students' behavior at the early stage of a semester, we construct a co-occurrence network from library check-in records to model students' social relationships. Figure 3 illustrates the constructed co-occurrence network partially. Each red node represents one student, while the edges between nodes indicate the co-occurrences of students when they check-in to the library. The width of the edges shows the number of co-occurrence time between them. Moreover, we use 5 times as the threshold to distinguish the "familiar strangers" and actual friends. The "familiar strangers" are the stranger students check-in to the library together by coincidence; so, the co-occurrence time between those students should be less than actual friends going to the library together. In the figure, solid black edges represent the co-occurrence times between nodes are higher than the threshold, while dashed gray edges represent the co-occurrence times between nodes are lower than the threshold. We model the learning behavior homophily among students by this co-occurrence network, which could further help the at-risk student prediction in the social feature layer. Because of social homophily, the features of students who have similar social connections should be close. We embed the co-occurrence network to encode the social homophily as representation vectors for every student by using Node2Vec [20]. Learning students' social homophily provides extra information for solving the data insufficiency challenges and enables EPARS to early predict STAR.

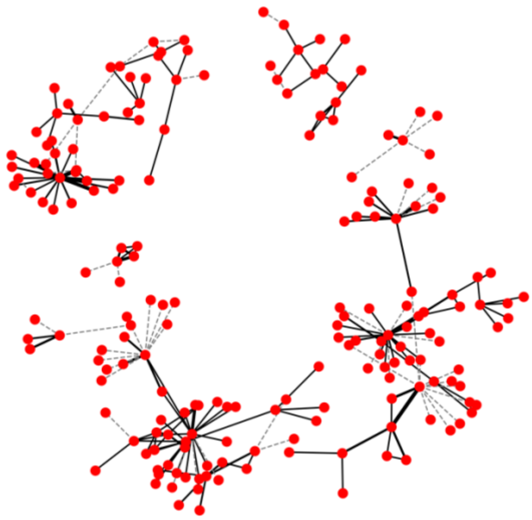

Fig. 3. A constructed co-occurrence network.

Last but not least, we augment the training samples of STAR by synthesizing new ones using random linear interpolation. After data augmentation, STAR have the same number of samples as that of normal students while training the classifier for prediction. It prevents the classifier from being dominated by the majority of normal students' samples which overcomes the challenge of extreme label imbalance in classification.

4.3 Results of Feature Fusion

We collect the data from 15,503 undergraduate students in an Asian university in 2016 to 2017 academic year for conducting experiments to validate the effectiveness of the

EPARS. There are 225 and 319 STAR in two semesters respectively. The experiment task is to predict STAR at the end of every week in the semester. The accuracy of STAR prediction (ACC-STAR) is defined as the amount of true positive predictions divided by the total number of STAR in the test set. The baseline approaches are handcrafted statistically significant behavior features (SF) and its combination with the components of EPARS including data augmentation (DA), regularity features (Reg), and social homophily features (SoH). All experiments are under 5-fold cross-validation and repeat 10 times. The results are reported in **Table 2** where the elements represent the average ACC-STAR. The proposed method outperforms all baselines from the first week to the end of the semester which confirms its effectiveness in STAR early prediction. Especially, our EPARS correctly predicts 61.84% STAR from their online and offline learning behaviors in the first week, which outperforms SF, DA, DA-Reg, and DA-SoH 38.22%, 17.50%, 14.62%, and 22.38%, respectively.

Table 2. Results of STAR early prediction.

Weeks	Baseline	DA	DA-SoH	DA-Reg	DA-Reg-SoH
1	0.447	0.526	0.505	0.539	**0.618**
2	0.395	0.421	0.447	0.618	**0.658**
3	0.395	0.408	0.461	0.539	**0.618**
4	0.308	0.368	0.447	0.592	**0.645**
5	0.408	0.421	0.513	0.592	**0.645**
6	0.447	0.447	0.539	0.566	**0.697**
7	0.395	0.500	0.461	0.605	**0.697**
8	0.539	0.421	0.474	0.632	**0.737**
9	0.572	0.408	0.441	0.592	**0.711**
10	0.487	0.444	0.487	0.671	**0.711**
11	0.539	0.582	0.574	0.671	**0.737**
12	0.500	0.582	0.595	0.684	**0.724**
13	0.539	0.608	0.618	0.684	**0.724**

5 Conclusion

In this paper, we propose a general research framework for learning analytics by extracting and fusing multilayer behavior which includes physical, social, and mental behavior. We demonstrate the feasibility and effectiveness of extracting social and mental features from textual data and spatiotemporal data. Also, feature-level fusion and model-level fusion methods reveal the flexibility of multilayer behavior fusion. According to the evaluation of automated online forum posts scoring and at-risk student early prediction, the proposed framework could effectively improve task performance.

Acknowledgements

The work is supported by Human-computer fusion cloud computing architecture and software definition method (project code: 2018YFB1004801). It is also supported by Learning Analytics and Educational Data Mining: Making Sense of Big Data in Education (project code: 1.61.xx.9A5V) and Multi-stage Big Data Analytics for Complex Systems: Methodologies and Applications (RGC No.: C5026-18G).

References

1. "Call for Papers of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)". Retrieved 12 February 2014.
2. Sin, K., & Muthu, L. (2015). APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS--A LITERATURE REVIEW. ICTACT journal on soft computing, 5(4).
3. Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Online Learning, 20(2), 13-29.
4. Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. EDUCAUSE review, 42(4), 40.
5. Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., & Kleinberg, J. (2010). Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences, 107(52), 22436-22441.
6. Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents' emotion regulation in daily life: Links to depressive symptoms and problem behavior. Child development, 74(6), 1869-1880.
7. Allen I E, Seaman J.: Changing course: Ten years of tracking online education inthe United States. In: Sloan Consortium. PO Box 1238, Newburyport, MA 01950 (2013)
8. Taghipour K, Ng H T.: A neural approach to automated essay scoring. In: Proceedings of the 2016 conference on empirical methods in natural language processing. 1882-1891(2016)
9. Sultan M A, Salazar C, Sumner T.: Fast and easy short answer grading with highaccuracy. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 1070-1075(2016)
10. Cox B, Cox B.: Developing interpersonal and group dynamics through asynchronousthreaded discussions: The use of discussion board in collaborative learning. In: Education, 128(4) (2008)
11. Yuan J, Kim C.: Guidelines for facilitating the development of learning communitiesin online courses. In: Journal of Computer Assisted Learning, 30(3): 220-232(2014)
12. Mikolov T, Karafiát M, Burget L, et al.: Recurrent neural network based language model. Eleventh annual conference of the international speech communication association. (2010)
13. Kalchbrenner N, Grefenstette E, Blunsom P.: A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188. (2014)
14. Richardson, V.: At-risk student intervention implementation guide. The Education and Economic Development Co ordinating Council At-Risk Student Committee p. 18 (2005)
15. He, J., Bailey, J., Rubinstein, B.I., Zhang, R.: Identifying at-risk students in massive open online courses. In: Twenty-Ninth AAAI Conference on Arti_cial Intelligence (2015)
16. Koprinska, I., Stretton, J., Yacef, K.: Students at risk: Detection and remediation. In: Proceedings of the 8th International Conference on Educational Data Mining. pp. 512-515 (2015)

17. Marbouti, F., Diefes-Dux, H.A., Madhavan, K.: Models for early prediction of at risk students in a course using standards-based grading. Computers and Education 103, 1-15 (2016)
18. Yao, H., Lian, D., Cao, Y., Wu, Y., Zhou, T.: Predicting academic performance for college students: A campus behavior perspective. ACM Transactions on Intelligent Systems and Technology (TIST) 10(3), 24 (2019)
19. Ellenbogen, S., Chamberland, C.: The peer relations of dropouts: a comparative study of at-risk and not at-risk youths. Journal of adolescence 20(4), 355-367 (1997)
20. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 855{864. ACM (2016)
21. Yu, Z., Du, H., Yi, F., Wang, Z., & Guo, B. (2019). Ten scientific problems in human behavior understanding. CCF Transactions on Pervasive Computing and Interaction, 1(1), 3-9.