Mentor Recommendation System

Jiwei Xiao
515021910592

May 2018

1 Introduction

The amount of data is experiencing explosive growth along with the progress of the times, students can search for mentor through search engines.The relevant information of the division is used as a reference, but this type of decision-making method still has information overload, incomplete information, and information asymmetry. And other issues.

Therefore, in order to help students make better decisions and achieve the optimization of two-way choices, they need to determine the key influence factors in the teacher-student relationship based on academic Big Data and Machine learning, design a comprehensive and intelligent tutor recommendation system based on the students' research interests and recommend suitable mentor for students.

2 Related work

Alarfaj, Kruschwitz et a1. (2012) designed a mentor recommendation system for enrolled PhD students, which combined data drive and database in two different ways. It recognized named entity in natural language processing to extract key words in the instructor's published paper text and self-labeling Keywords, in order to indicate their academic orientation characteristics.At the same time, students can input the keywords they are interested in, and the system will calculate the similarity between the keywords of the tutor and the students' search keywords to generate tutors for the measurement standards automatically.

Tehran University, Mosharraf, Taghiyareh (2012) mapped the problem of how students select their tutors into an optimization problem and proposed a corresponding solution based on genetic algorithms. The tutor and student's modeling information were derived from the Learning Management System (LMS),Community of Practice (CoP) and their own designed user survey system.

3 Methods

3.1 Teacher-student Relationship Analyisis

There are too much users' information that contains their paper cooperator and some details of their paper on our database. The teacher-student relationship network is shown as figure 1.In order to operate this useful imformation, we used SNAP [2] to store these materials.

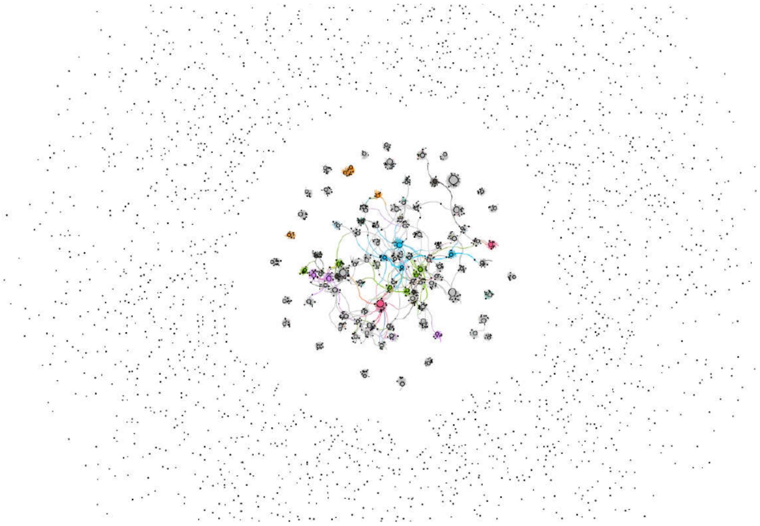


Figure 1: teacher-student relation network

SNAP is a open source library which is used for analyzing large social and information networks. Therefore, we use SNAP to handle our large amount of data. We have visualized the relasionship between our users well.

3.2 Features Extraction

As described above, we used SANP to transform our data to a large network. Then we need to extract the all imformation about two special users. We use node2vec [1] to improve ramdom walk algorithm. The main strategy is combining the BFS and DFS to improve the efficiency of searching. The model is shown as figure 2.

Then we chose thirty-two features as a feature vector between two users and input these feature vector into our model to predict the teacher-student relationship.

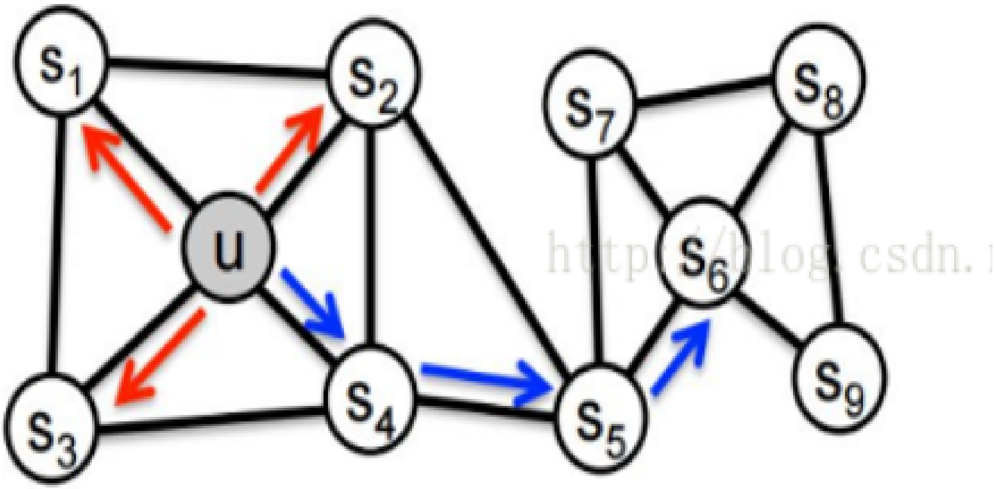


Figure 2: node2vec

3.3 Generate Negative Samples by PU algorithm

The public data source of the relationships between teachers and students is the website of AcademicTree.org and the tutor‘s personal homepage. We can build the teacher-student relationship data set by writing crawlers to get web contents. But we can’t get negative samples from any existing database. Therefore we used PU algorithm to generate negative samples from those positive samples.

Firstly, we used these positive samples to train a Bayesian network. Then we use the Bayesian model to evaluate all positive samples and regard the lowest score as threshold. Next we input some unlabel samples into our Bayesian model to get their score. Finally We regard them as negative samples if their scores are lower than that threshold

3.4 Train prediction model

After collecting these teacher-student relationship samples, we adopted multi-layer perceptrons, MLPs, the most frequently used artificial neural network algorithm, to deal with estimation and forecasting of hidden potential mentor relationships. We implemente MLP with Sequential Model from Python Keras, which proved to show better performance after we make comparison among several other models such as SVM and Decision-Tree. The model is shown as figure 3.

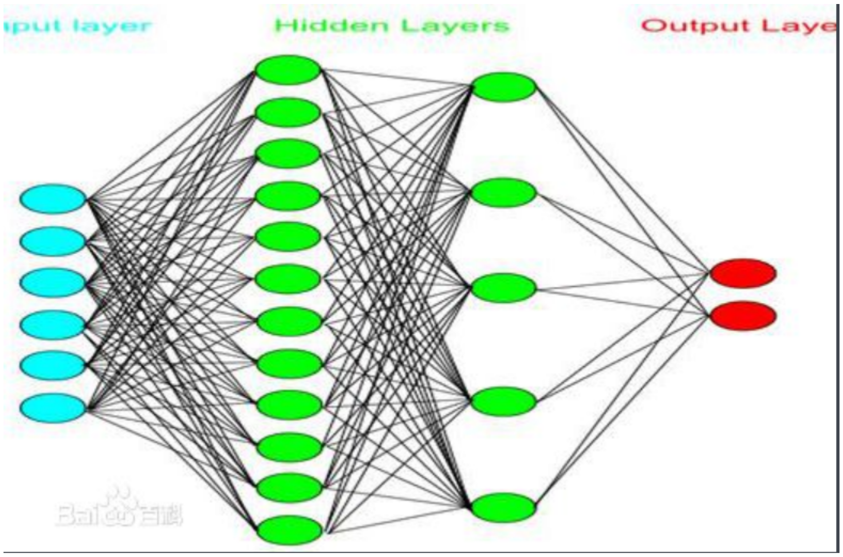


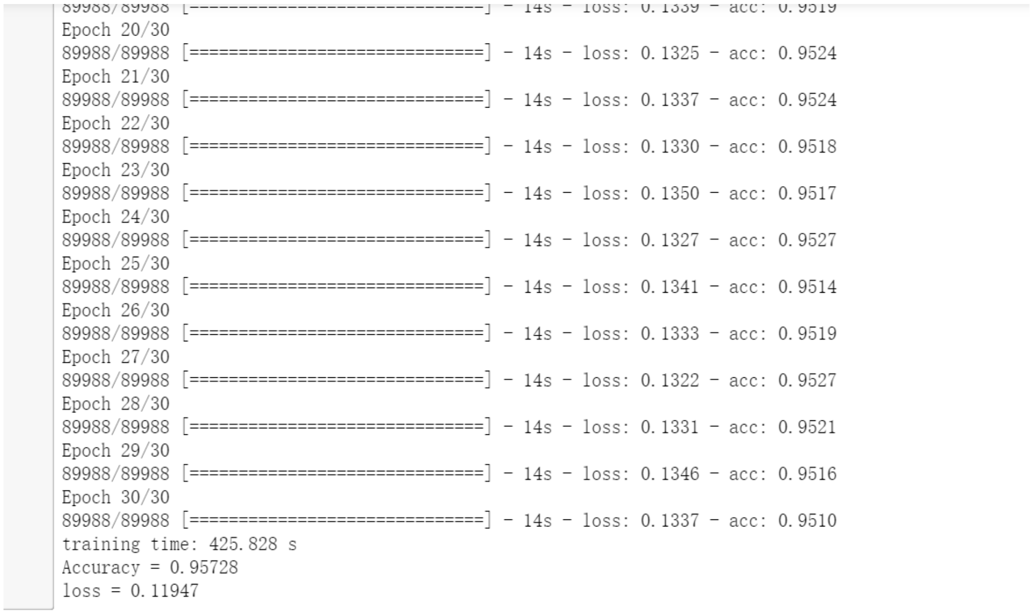
Figure 3: MLP

4 Experiment and Result

We used the Multi-layer Perceptron network to predict the teacher-student relationship. Our model gets good performance. The result of training process is shown as figure 4.

Then we used different dimensions of feature vector. The character relationship is shown as figure 5 and figure 6.

In the end, we tend to present this mentor relationship with our visualization system, which presents several kinds of map for user to understand the information of the mentor relationship network. The relationship network is shown as figure 7.



```
89988/89988 [==============================] - 14s - loss: 0.1339 - acc: 0.9519
Epoch 20/30
89988/89988 [==============================] - 14s - loss: 0.1325 - acc: 0.9524
Epoch 21/30
89988/89988 [==============================] - 14s - loss: 0.1337 - acc: 0.9524
Epoch 22/30
89988/89988 [==============================] - 14s - loss: 0.1330 - acc: 0.9518
Epoch 23/30
89988/89988 [==============================] - 14s - loss: 0.1350 - acc: 0.9517
Epoch 24/30
89988/89988 [==============================] - 14s - loss: 0.1327 - acc: 0.9527
Epoch 25/30
89988/89988 [==============================] - 14s - loss: 0.1341 - acc: 0.9514
Epoch 26/30
89988/89988 [==============================] - 14s - loss: 0.1333 - acc: 0.9519
Epoch 27/30
89988/89988 [==============================] - 14s - loss: 0.1322 - acc: 0.9527
Epoch 28/30
89988/89988 [==============================] - 14s - loss: 0.1331 - acc: 0.9521
Epoch 29/30
89988/89988 [==============================] - 14s - loss: 0.1346 - acc: 0.9516
Epoch 30/30
89988/89988 [==============================] - 14s - loss: 0.1337 - acc: 0.9510
training time: 425.828 s
Accuracy = 0.95728
loss = 0.11947
```

Figure 4: Training processing

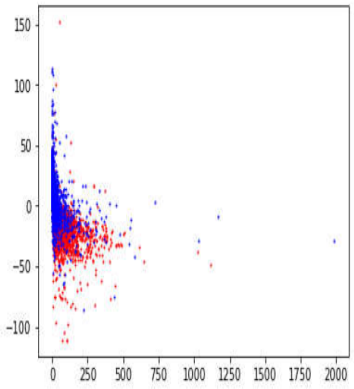


Figure 5: features 34

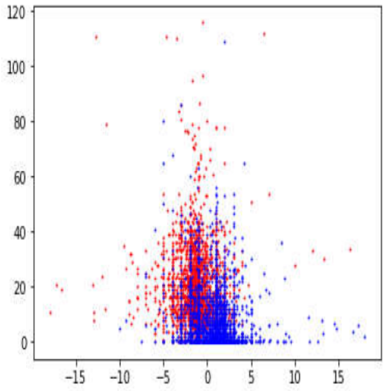


Figure 6: features 91

5 Conclusion

In this project, I have used many machine learning model and deepened my understanding of them. I have learned many ideas about how to deal with relationship prediction and social network.

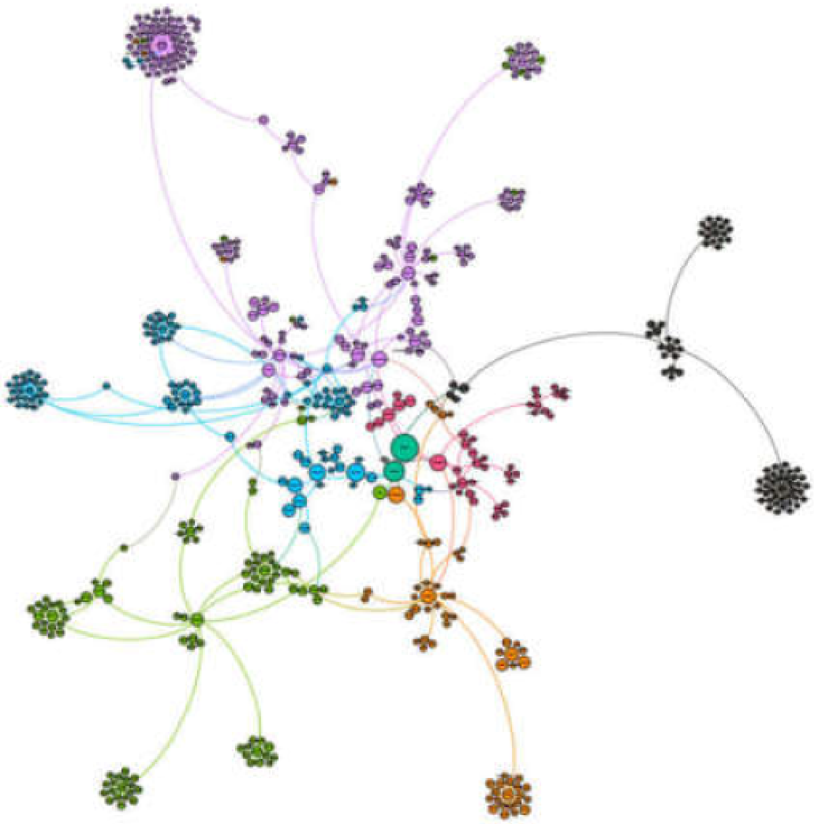


Figure 7: Visual relation network

References

[1] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *CoRR*, abs/1607.00653, 2016.

[2] Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1, 2016.