Career Prediction by Social Connection

陈彦锋 王远 毛适

Literature Review:

Companies such as Liepin, Linkelin and Pinterest have acheive great success in nowadays [2]. In other words, these kinds of companies do help a large group of people to find their suitable jobs, so the methodology behind these websites becomes researchers' interest. The conventional tools are build more based on people their career position, education and their previous jobs. Although it helps people are connected and find jobs with great satisfaction, this is still not the best methodology. The conventional tools neglect the other relationships of the users such as mentors and schoolmates. In this paper, we conduct machine learning algorithms and data mining to study the impact of social relations on the users' career change. More importantly, the application provides users a long term carrer plan instead of a short period of time.

To address our concern, we turn to a temporal network represetation of data and try to apply graph nerual networks. Recent researches includes generalizing the defination of convolution to noneucclidean space to directly achieve graph convolutional network, learning a good embedding that representing network nodes as low dimensional vectors[1].

Motivation:

In recent years, more and more attention has been paid to predicting the changes of user positions based on user career trajectory. In general, the conventional method is based on the user's resume to extract the characteristics of the career trajectory, such as company, position, education, time, position accumulation and other characteristics of position change, to analyze the influencing factors of position change through historical trajectory data. More significantly, the data can be used to train to predict the next position for which user would undertake in the future. However, the existing methods of career trajectory prediction often neglect the social relationships between users, such as colleagues, mentors and so on.

This research will combine the user's career trajectory and the characteristics of their nearest neighbors in social networks to conduct data mining in order to study the possible impact of social relations on the user's career change. This will obtain a higher accuracy of career trajectory prediction.

Methodology:

1. Extracting the resume of users from the government website (http://www.sz.gov.cn/cn/xxgk/zfxxgj/sldzc/sz_97404/crg/) by using web crowler.
2. Parsing information into more detail based on postion location, position time, etc. During this process, the knowledge of naturual language processing (NLP) will be used.
3. Building tempral network representation based on the parsing information and using graph neural network to learn useful feature embeddings, which preserves both network topology structure and note content infromation, for the downstream predictor. Our attempts focus more on learning a good embedding followed by traditional machine learning algorthms sides, but will also try end-to-end maner to perform the utimate prediction directly. However, consider the temporal charatctorization of our data, proper methods are needed
4. Using different kinds of machine learning algorithms to perform training on the data we extracted to predict the next position of a particular users will undertake in future. In particular, the analysing differetnt factors affect users to change workplaces.

Intended experiment:

We intend to split our collected data as training and testing sets. To evaluate the performance of our algorithm, we first train our model on the training set where the data from all the time periods collected are given. Then, we apply the trained model to testing set, where the data from the final period of time are intentially blocked and serves as ground truth of predicting results. However, the method to measure the similarity of the results and ground truth needs to be defined. We plan to divide the whole career set into several different levels and try to predict which levels the user will enter. Then, we will evaluate our model's performance by calculating the accuracy of the prediction. If the test data prediction is in the right level, we will count it as a correct condition.

Reference:

1. Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. arXiv preprint arXiv:1901.00596, 2019.
2. https://lmtw.com/mzw/content/detail/id/155422/keyword_id/-1