MATH 6380J Final Project: ELO Rating and Network Link Prediction on College Ranking Data

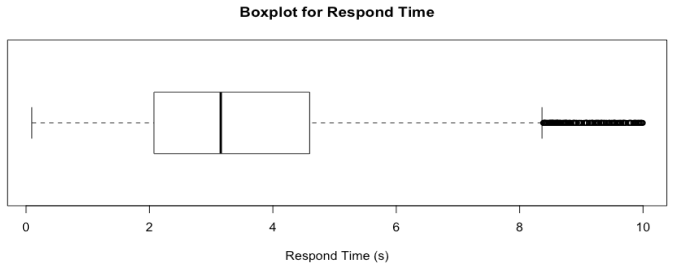
Chenyang Dong[1], Tsz Cheung LO[2] and Jiacheng Xia[2] {cdongac, tcloaa, jxiaab}@connect.ust.hk
[1]: Department of Mathematics, HKUST [2]: Department of Computer Science and Engineering, HKUST

1. Introduction

In this project, we focus on the analysis of pairwise comparison data. Two approaches are involved. The first is Elo system, which starts from equal score and re-estimates school's level after a popularity "competition" between two schools are made. The second is to evaluate the importance of a node using "status", a recent hot topic in computer science especially social network analysis, that a in-degree to degree ratio models the importance of the data. We got an overall accuracy in 70-75% range for both methods, and discussed what could have hindered us from getting better results.

2. The World College Ranking Dataset

The dataset consists of 8823 pairwise comparisons (7470 valid ones) of 261 colleges. The college ID's are indexed from 293152 to 293416 (four missing ID's within this range). The randomly selected pairwise comparisons are evaluated by 340 different annotators. For each comparison, the annotator's response time (seconds), the winner and loser college ID are recorded in the dataset. Other attributes are not studied in this project. Notice the distribution of response time is highly skewed to the right, we remove the instances with response time larger than 10 seconds, which are considered to be outliers statistically according to the boxplot.



3. Elo-rating system

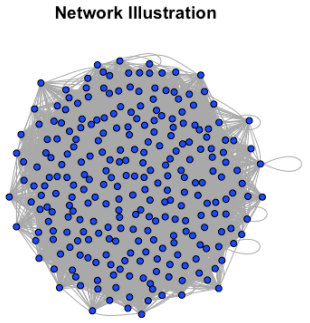
Elo-rating system is widely adopted to estimate a player's strength in gaming systems.
r'(1) = r(1) + K * (S(1) – E(1)); r'(2) = r(2) + K * (S(2) – E(2))
Where r(1), r'(1) are old and new score in each iteration, and S(1) is the observed game result, 1 for the player wins. E(1) is the estimated probability of player 1 wins.
Examples adopting this system include FIFA, CodeForces, etc. We set initial score to be 1000 and step size to be k=10. Using a 10-fold cross validation after random shuffling, we observe that the average score is approximately 0.7.

4. Network Construction and Link Prediction

Based on the status theory, the college ranking network is expected to have the acyclic property approximately. More specifically, if there exists a path from node A to node B, then the status of B is higher than that of A. For each node, a larger in-degree and lower out-degree jointly imply its relative prestige.

$$prestige(v) = \frac{in\_degree(v)}{in\_degree(v) + out\_degree(v)}$$

Directed networks were constructed from the training data by 10-fold Cross Validation. Then the distances between each pair were computed through Floyd-Warshall algorithm. The prestige value for each node was computed by incorporating in-degree and out-degree information of neighbors. We predicted the link direction between two universities based on both the path accessibility and the relative prestige in each testing case. An overall average testing accuracy of more than 73% could be achieved.



5. Analysis on respond time

We tried to modify the impact on K based on a user's respond time. However we found that this doesn't make much difference observed from cross-validation scores. Naively using a penalty based on users respond time, or cutting too-fast and too-slow responses do not have good effect in our experiments. We therefore conclude that, due to abnormal cases' existence, it is probably very hard to modify the algorithm so as to get a better grade.

6. Conclusion

In this project, we used both iterative-tuning methods and status method from network theory, and achieved a prediction accuracy ranging from 70~75%. Notice that abnormal input data still exists in the processed dataset (e.g. extreme values of the response time), and there exist significant bias in annotators' decisions. We conclude that without proper anomaly detecting method, this is probably best accuracy one can get. One may improve the prediction accuracy by classifying the decision-makers based on their behavior and eliminating "noisy" ones.

7. Acknowledgement

We appreciate prof. Yuan Yao for the excellent lectures throughout the semester, and we welcome all kinds of feedbacks on this project.

8. References

[1]. Qianqian Xu, Jiechao Xiong, Xiaochun Cao, and Yuan Yao. False Discovery Rate Control and Statistical Quality Assessment of Annotators in Crowdsourced Ranking, ICML 2016
[2]. Bourdieu, Pierre. Distinction: a Social Critique of the Judgment of Taste, translated by Richard Nice. *Cambridge: Harvard University Press,* 1984.
[3]. Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg. "Signed networks in social media." *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010.