MATH 6380J Final Project: Rank Aggregation Models on World College 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

Introduction

Rank Aggregation Problem

In short, Rank Aggregation Problem is that, given many “local” comparison results, how can you output an optimal model that can be use for predict “future” comparison, or, generate a global ranking for reference.

Project Overview

In this final project, we focus on the analysis of pairwise comparison data. **2 models** are involved. The first is **Elo Rating 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 around 70-75%** for both methods, and discussed what could have hindered us from getting better results.

World College Dataset

- 8823 pairwise comparisons from 340 participants
- 261 colleges (Harvard, Yale, CMU, Princeton, HKUST etc.)

Winner ID	Winner Text	Loser ID	Loser Text	Response Time (s)
293212	Pennsylvania State University, USA	293377	Beijing Institute of Technology, China	8.514
293273	University of California, Santa Cruz, USA	293195	Washington University in St Louis, USA	4.763
293156	Massachusetts Institute of Technology, US/	293308	Texas A&M University, USA	6.282
293155	Harvard University, USA	293192	New York University, USA	2.806
293245	Michigan State University, USA	293373	Central South University, China	10.544
293167	Johns Hopkins University, USA	293326	Rensselaer Polytechnic Institute, USA	3.085

Figure. Snapshot of Data

1	Yale University
2	Harvard University
3	Cornell University
4	Princeton University
5	Stanford University
6	University of California
7	University of Cambridge
8	University of California
9	University of Oxford
10	California Institute of Technol

Figure. Expect to output a global ranking

Model 1: Elo Rating System

Model Introduction

Elo-rating system is widely adopted to estimate a player’s strength in gaming systems.

$$r'(1) = r(1) + \big(S(1) - E(1)\big)\,K = r(2) + \big(S(2) - E(2)\big)\,K$$

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 Using Elo

Elo Rating System is used in **FIFA, CodeForces** (algorithm competition), etc.

Rank	Name	♂ ♀	Flag	Elo
1	Ke Jie	♂		3620
2	Park Junghwan	♂		3593
3	Mi Yuting	♂		3571

Figure. World Go (围棋) Ranking

Accuracy of College Data

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**.

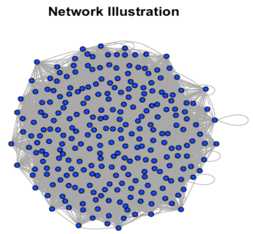
Model 2: Network Construction and Link Prediction

Model Introduction

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.



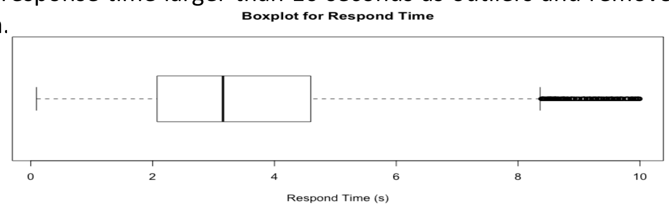
Accuracy

An overall average testing accuracy of **more than 73%** could be achieved.

Conclusion

Does Time Response Matter?

The distribution of participant’s response time (seconds) for each comparison, is highly skewed to the right, so we treat instances with response time larger than 10 seconds as outliers and remove them.



But abnormal input data still exists in the cleaned dataset (e.g. extreme response time), and there exists significant bias in annotators’ decisions. We conclude that without proper anomaly detecting method, this is probably best accuracy (**70-75%** for both iterative-tuning methods and status method) we can get. One may improve the prediction accuracy by classifying the decision-makers based on their behavior and eliminating “noisy” ones.

[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.