Final Project
World College Ranking with Pairwise Comparisons

RONG Yi
Department of Mathematics
yrong@connect.ust.hk

LIU Haoyu
Department of Information Systems, Business Statistics and Operations Management
hliuaw@connect.ust.hk

Abstract

In the project, we make use of a crowdsourcing dataset on pairwise comparisons to rank world colleges. By applying Hodge theory to the statistical ranking problem, we achieve reasonable world college ranking, based on the dataset. We compare the results generated by different generalized linear models, as well as the results on `allourideas`. We also present some further analysis and extensions of the project.

1 Introduction

Collecting data from Internet users is becoming increasingly fast, easy, and more importantly, inexpensive, because of the maturity of crowdsourcing technologies. Compared with traditional survey and experiment methods, crowdsourcing manages to obtain observations from a much more general masses rather than a limited coterie. However, every rose has its thorns. The participants are from such different and distributive backgrounds that having control of the implementation is of great difficulty.

The dataset we use for the world college ranking project is from crowdsourcing platform, which consists of 261 candidate colleges and 340 distinct voters from different districts. The voters are randomly shown with two colleges each time and asked to choose which college they would rather attend or cannot decide. There are totally 8823 pairwise comparisons collected.

Mathematically, we applied Hodge theory to the statistical ranking problem. Hodge decomposition of pairwise comparisons may achieve reasonable global scores, even though the dataset is incomplete or imbalanced. It can also be computed easily by least squares.

2 Algorithm

The algorithm is adapted from the paper [3].

After initialization, we compute the gradient δ_0 and the curl δ_1. The conjugate of δ_0 is

$$\delta_0^* = \delta_0^T * diag(W)$$

The unnormalized graph Laplacian can be computed as

$$\Delta_0 = \delta_0^* * \delta_0$$

The divergence operator can be computed as

$$div = \delta_0^* * \hat{Y}$$

Finally, the global score is

$$\hat{s} = lsqr(\Delta_0, div)$$

We sort the scores in descending order and obtain the ranking. Furthermore, we may compute the first, second, and third projections on gradient, harmonic, and curl flows respectively.

The Hodge decomposition decomposes the pairwise comparison flows $\hat{Y}$ into three parts.

$$\hat{Y} = \hat{Y}^g + \hat{Y}^h + \hat{Y}^c$$

They are gradient flows $\hat{Y}^g$ (corresponding to globally acyclic), harmonic flows $\hat{Y}^h$ (corresponding to locally acyclic), and curl flows $\hat{Y}^c$ (corresponding to locally cyclic), respectively. The first two parts are locally consistent (which is curl free), while the last two parts are inconsistent (which is divergence free). The paper [1] contains detailed theoretical development of Hodge decomposition of pairwise comparisons.

3 Results

We have tried several generalized linear models, such as uniform model

$$\hat{Y}_{ij} = 2\hat{\pi}_{ij} - 1$$

Table 1: World College Ranking (Top 10, Uniform Model)

Ranking	University	Score
1	Yale University, USA'	0.882455019
2	Harvard University, USA	0.876313478
3	Princeton University, USA	0.844395535
4	Cornell University, USA	0.831344329
5	University of Cambridge, UK	0.805034486
6	Stanford University, USA	0.802202174
7	University of California, Los Angeles, USA	0.782118932
8	University of California, Berkeley, USA	0.778413105
9	University of Oxford, UK	0.732455321
10	California Institute of Technology, USA	0.709093668

Bradley Terry model

$$\hat{Y}_{ij} = \log \frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}$$

and angular transform model

$$\hat{Y}_{ij} = \arcsin(2\hat{\pi}_{ij} - 1)$$

We present the results for all generalized linear models mentioned above and compare them. Also, we compare our results with the results on `allourideas`. The result on `allourideas` is as of May 21, 2017, which is available at the link

`allourideas.org/worldcollege/results`

We find our results are extremely close to the results on `allourideas` (Table 4), which suggests that our results are reasonable.

Specifically, the ranking generated by the uniform model (Table 1) and the ranking generated by the angular transform model (Table 3) have great similarity, though they have different scores. When we look at the Top 10 universities, the only differences appears at University of Cambridge, UK (ranking 5 in ours while ranking 11 on `allourideas`) and Massachusetts Institute of Technology, USA (ranking 13 in ours while ranking 5 on `allourideas`).

Moreover, the ranking generated by the Bradley-Terry model (Table 2) also makes sense, but are not the same as the previously mentioned ones. Specifically, the Top 10 list includes universities such as Imperial College London, UK (ranking 8 in ours while ranking 29 on `allourideas`) and University of Chicago, USA (ranking 10 in our while ranking 18 on `allourideas`). However, we cannot deny that they are all strong candidates for the Top 10 list, given their fame and reputation.

Table 2: World College Ranking (Top 10, Bradley Terry Model)

Ranking	University
1	California Institute of Technology, USA
2	University of Oxford, UK
3	Stanford University, USA
4	Harvard University, USA
5	Massachusetts Institute of Technology, USA
6	Princeton University, USA
7	University of Cambridge, UK
8	Imperial College London, UK
9	University of California, Berkeley, USA
10	University of Chicago, USA

Table 3: World College Ranking (Top 10, Angular Transform Model)

Ranking	University	Score
1	Yale University, USA	1.386490732
2	Harvard University, USA	1.377209998
3	Princeton University, USA	1.326891804
4	Cornell University, USA	1.306147288
5	University of Cambridge, UK	1.264477022
6	Stanford University, USA	1.260033202
7	University of California, Los Angeles, USA	1.228551079
8	University of California, Berkeley, USA	1.222050557
9	University of Oxford, UK	1.149885667
10	California Institute of Technology, USA	1.113239048

4 Limitations

The size of the pairwise comparison matrix is 261×261, but we have only 8823 pairwise comparisons. Thus, the pairwise comparison matrix is seriously sparse. It costs us a lot of computational works to compute Hodge decomposition of three orthogonal components, that is, gradient flows (globally acyclic), harmonic flows (locally acyclic), and curl flows (locally cyclic). Thus, we do not distinguish local and global inconsistency here and combine the divergence free inconsistency terms together.

Table 4: World College Ranking on `allourideas` (Top 10)

Ranking	University	Score
1	Princeton University, USA	92
2	Harvard University, USA	91
3	Stanford University, USA	90
4	University of California, Berkeley, USA	90
5	Massachusetts Institute of Technology, USA	88
6	University of California, Los Angeles, USA	88
7	Cornell University, USA	88
8	Yale University, USA	87
9	California Institute of Technology, USA	84
10	University of Oxford, UK	84

Acknowledgments

The authors thank Professor YAO Yuan for the course MATH 6380, the knowledge of applied Hodge theory and the dataset

`math.stanford.edu/~yuany/course/data/college.csv`

Reference

[1] Jiang, Xiaoye & Lim, Lek-Heng & Yao, Yuan & Ye, Yinyu (2008). Statistical ranking and combinatorial Hodge theory. *arXiv:0811.1067.*

[2] Xu, Qianqian & Xiong, Jiechao & Cao, Xiaochun & Yao, Yuan (2016) False Discovery Rate Control and Statistical Quality Assessment of Annotators in Crowdsourced Ranking. *arXiv:1605.05860.*

[3] Xu, Qianqian & Huang, Qingming & Jiang, Tingting & Yan, Bowei & Lin, Weisi & Yao, Yuan (2012). HodgeRank on Random Graphs for Subjective Video Quality Assessment. *IEEE Transactions on Multimedia, 14(3), 844-857.*