World College Ranking

Group memeber: Ye Rougang, Tan Chunxi, Han Ruijian

May 21, 2017

1 INTRODUCTION

The dataset we choose is the worldcollege ranking, which is generated from a survey about choosing one better university from a pair of options each time. The survey is built on the website named Allourideas crowdsourcing platform. There are nearly 8300 pairwise comparison results for 261 colleges in the dataset. Each one recoded the winner and the loser college.

As using pairwise comparisons to rank items is one of key problems in machine learning, which has a lot of applications related to ranking issues, such as recommending products, advertisement personalization and matching making. It's demonstrated that pairwise comparison queries take shorter time to be responded compared with other query types like likert scale (which often show as 5 level likert scale: strongly dislike, dislike, indifference, like, strongly like) etc.

Yet in many cases, for the problem related to n items we usually cannot collect $n(n-1)/2$ comparisons, that is, we cannot get all of pairwise comparisons for different items. It can be regarded as the incomplete situation. Here we try to use different methods to get the ranking of different colleges.

In this project, we used three different ways to get the ranking. The first approach is the traditional model named Bradley-Terry-Luce model (BTL model). The second way is a Markov chain approach named rank centrality. The last method we use is baesd on the TrueSkill ranking system which is put forward by Microsoft.

2 THE BRADLEY-TERRY-LUCE MODEL

2.1 BTL MODEL

In this section, we use the BTL model to get the ranking of these colleges.
Consider the basic case of two options, where we let the Gaussian random variables I and J represent the quality of option i and option j respectively, i.e.

$$I \sim N(\mu_i, \sigma_i^2), J \sim N(\mu_j, \sigma_j^2).$$

Then we can firstly focus on the mean values to decide the ranking. When $\mu_i > \mu_j$, we can regard option i ranks higher than option j.

The Bradley-Terry-Luce model (BTL) gives a estimate for the difference of mean values of two options, which is

$$\hat{\mu}_{ij} = ln(\frac{A_{i,j}}{A_{i,j} + A_{j,i}}) - ln(1 - \frac{A_{i,j}}{A_{i,j} + A_{j,i}})$$

where $\hat{\mu}_{ij}$ here is the estimate of $\mu_{ij} = \mu_i - \mu_j$, $A_{i,j}$ is the empirical counting of comparisons of choosing option i over option j.

To determine the quality scores (mean values here) for m options, denote the vector of quality scores $\mu = [\mu_1, \mu_2, \cdots \mu_m]$. Let D be a $m \times m$ matrix which forms by $\hat{\mu}_{ij}$, then the least squares estimate is gotten by minimizing the squared error between estimates and true values:

$$\hat{\mu} = \underset{\mu \in \Re^m}{\arg\min} \sum_{i,j} (D_{i,j} - (\mu_i - \mu_j))^2$$

One common approach is to assume that the mean of all the $\hat{\mu}_i$ is zero. In this case, the least squares solution is

$$\hat{\mu}_j = \sum_{i=1}^{m} \frac{D_{i,j}}{m}$$

2.2 RESULT AND ANALYSIS

In our dataset, $m = 261$, and I, J indicates two different compared colleges. Applying the above LSE solution of the BTL model with the Python, a vector of estimate values μ can be collected and briefly gives a ranking of these colleges. There shows a part of results(the top 10 universities in the ranking of dataset).

Table 2.1 gives a detailed ranking with values, of which the graph is shown in Figure 2.1. It seems reasonable at the first glance, yet we notice in Figure 2.1 there is a sharp jump of ranking values between the third and the fourth one. This might be avoided for a much bigger dataset.

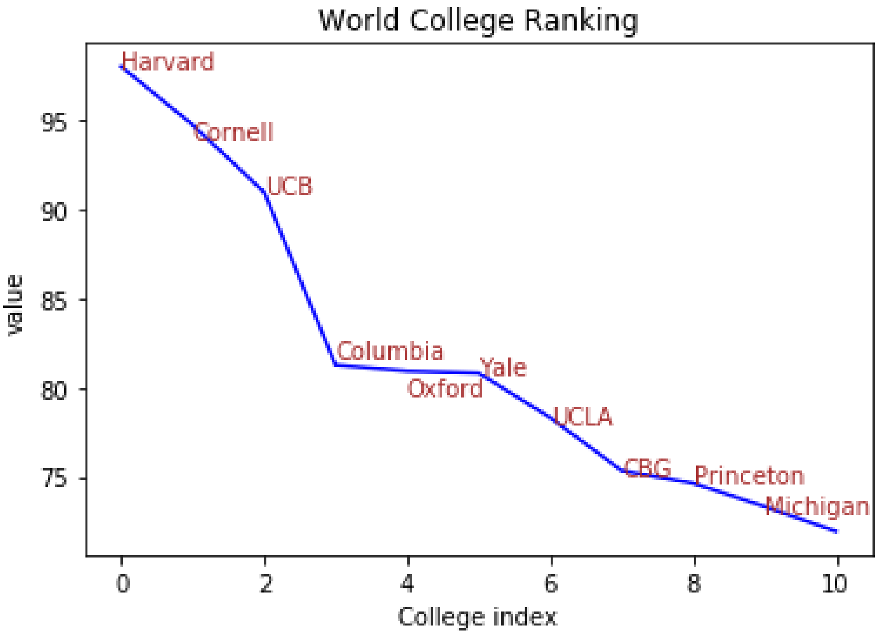

Figure 2.1: Trend of Top 10 Universities by BLT

College	Value	Rank
Harvard University, USA	98.08	1
Cornell University, USA	94.79	2
University of California, Berkeley, USA	91.01	3
Columbia University, USA	81.27	4
University of Oxford, Uk	80.94	5
Yale University, USA	80.82	6
University of California, Los Angeles, USA	78.31	7
University of Cambridge, UK	75.33	8
Princeton University, USA	74.62	9
University of Michigan, USA	73.31	10

Table 2.1: Top 10 Universities in the world by BTL

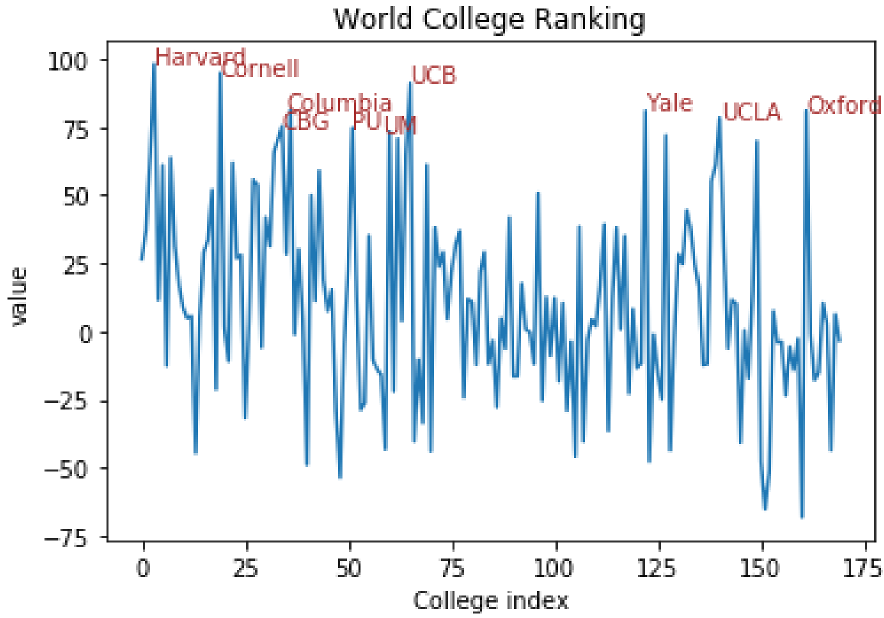

Figure 2.2: The Ranking of All Universities in the world by BTL

In the Figure 2.2, more information could be figured out. For instance, the difference of value between top universities and relatively low ranked universities is high. The top one is Harvard University of around 98 in value and the lowest one is around -68.

3 Rank Centrality

3.1 Markov Chain approach

To some extent, the rank centrality is based on the BTL model. In this approach, we regard the colleges as nodes in one graph. Then we define a random walk on the graph. A brief graphic model is shown in the figure 3.1.

By using the dataset to build the transition matrix. The transition probability is computed as follows:
Let A_{ij} be the fraction of times that object j has been preferred to object i. So

$$P_{ij} = \begin{cases} \frac{1}{d} A_{ij}, & i \neq j \\ 1 - \frac{1}{d} \sum_{k \neq i} A_{ik}, & i = j \end{cases} \tag{3.1}$$

where d is the maximum degree of a node. This ensures that each row-sum is at most 1. Then, in order to ensure that each row-sum is exactly one, we add a self-drop to each node. So we get the transition probability above. And then we compute the limiting distribution

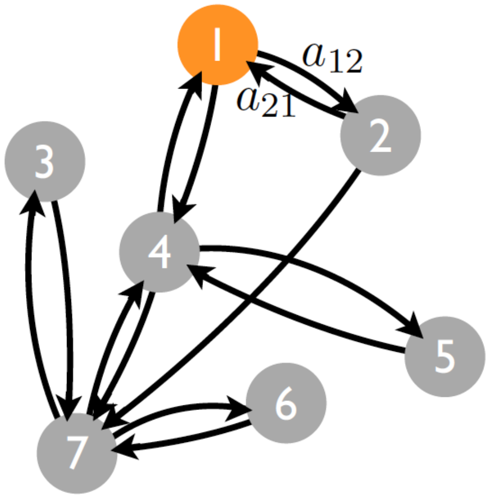

Figure 3.1: A Brief Random Walk Graphic Model

for each node.

$$\pi P = \pi$$

In our question, for one node, the higher value its limiting probability is, the higher rank it has.This model is come up with by Sahand Negahban, Devavrat shah and Sewoong Oh in the paper named "Iterative Ranking from Pair-wise Comparisons". They proved this method work well when the number of pairwise comparisons is at least $O(n \log n)$ for n items. In our dataset, $n = 261$, the number of pairwise comparisons is nearly 8300, which is larger than $n \log n$. Thus the method should perform well.

3.2 Result and Analysis

We show our results in the table 3.1, figure 3.2 and table 3.2.

In the two tables, we list the ranking for the top 10 universities in the world and the top 10 in China. The column named "Value" stands for limiting probability for each node(college). Here, we did not normalize values, since normalization have no effects on the ranking. The column named "Rank" shows the ranking for these 261 universities. In the survey, we can see users are more likely to select Harvard University as the best university. But it just stands for some of people, it is no doubt that the result will be more precise if our dataset is bigger.

In the table 3.1, the result is reasonable, but the result of top 10 universities in China in the table 3.2 is interesting. Actually in our comparison dataset, universities of China is always losers when comparing to foreign universities, except for few universities. If one university occurs more frequently, the fraction of defeated times should be larger. So it is really hard to rank universities which have low ranks from the dataset. This method works well to rank

College	Value	Rank
Harvard University, USA	0.402	1
Stanford University, USA	0.319	2
Yale University, USA	0.301	3
Cornell University, USA	0.274	4
Princeton University, USA	0.264	5
University of California, Los Angeles, USA	0.240	6
University of Cambridge, UK	0.216	7
University of California, Berkeley, USA	0.204	8
Massachusetts Institute of Technology, USA	0.169	9
California Institute of Technology, USA	0.138	10

Table 3.1: Top 10 Universities in the world

College	Value	Rank
Peking University, China	0.110	15
Tsinghua University, China	0.052	41
Shanghai Jiaotong University, China	0.037	55
Fudan University, China	0.028	80
Zhejiang University, China	0.027	83
Tongji University, China	0.026	87
Nanjing University, China	0.025	95
Renmin University of China	0.024	99
University of Electronic Science and Technology, China	0.023	105
Beijing Normal University, China	0.022	107

Table 3.2: Top 10 Universities in China

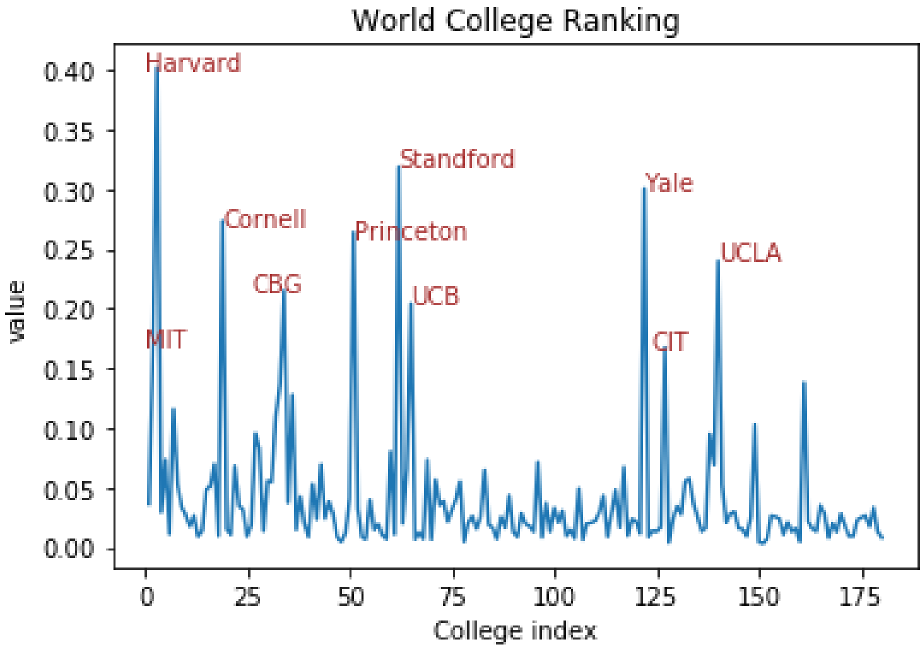

Figure 3.2: The Ranking of All Universities by Rank Centrality

universities which always win, but may not fit for universities which always lose. Although this method may not work so well in ranking top 10 universities in China, we will get better result if we change our survey including more pairwise comparisons among the universities of China.

4 TrueSkill Rating Sytem

As the pairwise comparisons is like a head-to-head match between two players in a competitive game. We can seek some inspirations from skill rating systems and to apply them to regard different colleges as various players to get corresponding ranking.

There exist many notable skill rating systems such as *ELo system* which was adopted by the World Chess Federation in 1970. Here we would use another remarkable rating system called *TrueSkill* that is developed by Microsoft for both matchmaking and player ranking for the online gaming system. It can model multiple teams game as well as free-for-all modes. We use this ranking system to head-to-head mode to update skill ratings of these colleges after each pairwise comparisons.

As *TrueSkill* is a Bayesian graph algorithm, we assume each college owns a Gaussian prior rating $s \sim N(\mu, \sigma^2)$. Then from Bayes' rule, we could update ratings given comparison results, i.e.

$$p(s|r) \propto p(r|s)p(s)$$

College	Distribution	Rank
Yale University, USA	$N(37.25, 1.61^2)$	1
Cornell University, USA	$N(35.47, 1.47^2)$	2
Harvard University, USA	$N(35.03, 1.63^2)$	3
Princeton University, USA	$N(34.62, 1.55^2)$	4
Stanford University, USA	$N(34.39, 1.51^2)$	5
Tsinghua University, China	$N(28.43, 1.15^2)$	38
Hong Kong Polytechnic University	$N(26.13, 1.16^2)$	73
National Taiwan University, Taiwan	$N(24.70, 1.21^2)$	116
University of Southampton, UK	$N(22.81, 1.28^2)$	166
Hunan University, China	$N(20.11, 1.22^2)$	233

Table 4.1: Ratings of Some Chosen Colleges (ranking is given by μ values)

where r refers to the comparison result. By using the factor graph and the approximate message passing based on the sum-product algorithm, one can infer the posterior distribution, which is exactly the updated skill rating $s^{new} \sim N(\mu^{new}, \sigma^{new2})$after pairwise comparisons. μ value follows a college's win/lose record. Higher value means higher rating. While σ value follows the number of comparisons involved. Lower value means more comparisons involved and higher rating confidence. Thus the *TrueSkill* can reflect uncertainty to the rating result and works well in this incomplete dataset.

4.1 Result and analysis

After the python coding and given the initial rating followed $N(25, \frac{25}{3}^2)$, we got the rating for each college, and we chose ratings of some colleges as shown in table 4.1. The μ value changes more quickly for higher ranking universities compared with middle ranking universities. The μ value decrease up to 3 from the 1st to the 5th college while from Tsinghua University(38th) to Hong Kong Polytechnic University(73rd), it only changes by 2.3. This can be explained that the middle ranking colleges usually lose with the higher ranking colleges and win with the lower ranking ones which leads to closer ratings.

Further, the σ value is significantly higher for those top and bottom universities and lower for the middle ranking ones. This should be related to the survey designing that makes comparisons of middle ranking items show more often to estimate their positions more accurately. These observations can be visualized through figure 4.1. And figure 6.1 displays normal curves of top 5 colleges by *TrueSkill* rating system which one can see it's a vivid and informative representation of ratings of these colleges.

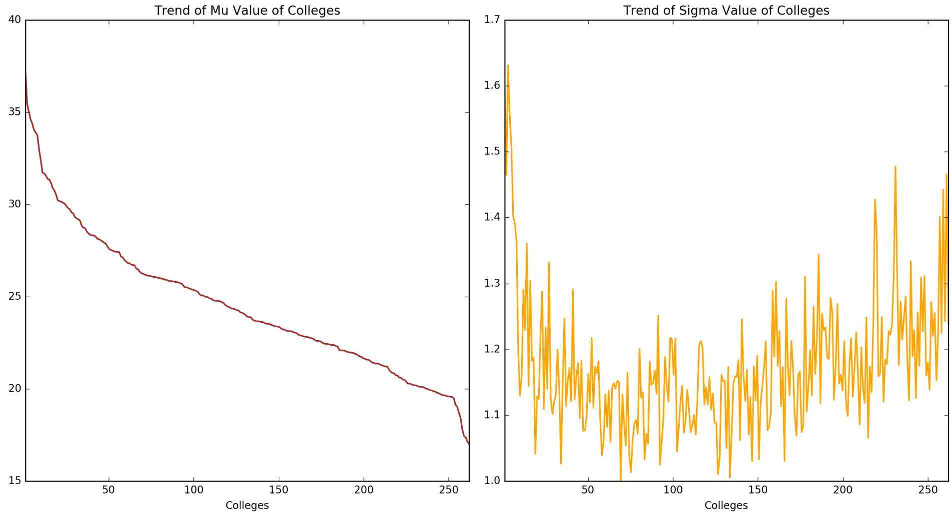

Figure 4.1: Trend of μ and σ of Ranking of Colleges by TrueSkill

5 SUMMARY

For this world college ranking dataset which mainly consists of pairwise comparison results, we start from the classical Bradley-Terry-Luce model and use the least square solution to get the ranking. Then use the Markov Chain approach to get limiting distribution as the judgment of ranking. Finally, we use the idea of player ranking in competitive games to summarize more detailed ranking of these colleges. The later two method are good choices when facing incomplete pairwise comparison data.

And actually by *TrueSkill*, we can get information about the comparison quality which can be regarded as a kind of expression of the draw probability in one comparison. Noticing there contains response time in the dataset, we think that by transforming it to be the draw probability in each comparison, we can predict more precisely when facing two options (including a option meaning *draw* in this case) as the longer response time, the harder a choice can be made between the two.

6 REMARK OF CONTRIBUTION

Han Ruijian preprocessed the dataset and used the Markov Chain method to get the ranking.

Tan Chunxi used the BTL model to transform the problem and get the ranking by the LSE solution.

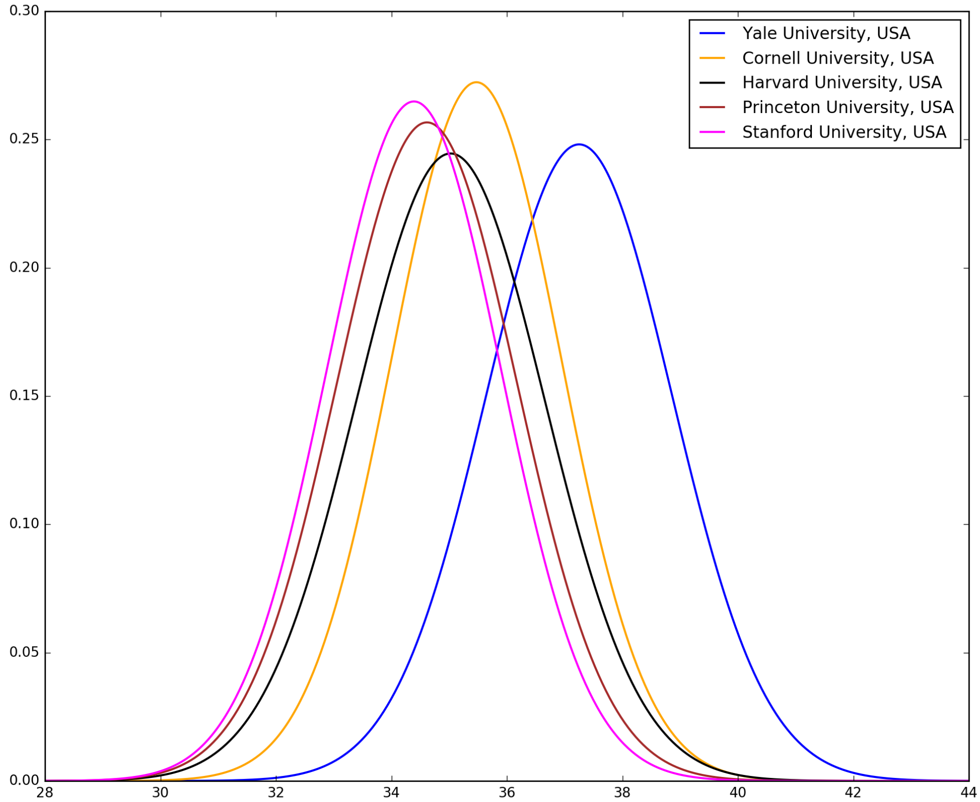

Figure 6.1: The Normal Curves of Top 5 Universities by TrueSkill

Ye Rougang used the *TrueSkill* rating system to model the ranking and summarized results of different methods.

Codes and ranking results (saved as csv files) can be found in the attachments.