ANALYSIS OF PAIRWISE VOTINGS ON PROJECT 2

{ ESTHER VISSER } HKUST

INTRODUCTION

In this project I investigated different methods to define a ranking for the given pairwise votings. In the end I want to find out the similarities of the results of all the methods. The methods that were used are Borda Count, kennemy optimazation, using a random walk and lasty I determined the Concordet winner.
Every ranking that will be given in this poster is a list of numbers. These numbers correspond to the numbers given in the name of the posters. For example, the first poster is named: 01.Bu, Qi <hyperlink>. This will be denoted by the number 1 in the rankings.

RANDOM WALK

As described in [3][2] we can define a graph on the set of posters (1 up and until 12) where we have an edge if the two posters are compared. Then we can define the probability of going from poster i to a different poster j as $P_{ij} = \frac{a(i,j)}{a(i,j)+a(j,i)}$ assuming there is an edge between them, where $a(i,j)$ is the same as the $a(i,j)$ used in the Concordet method: it measures how many times j wins of i. If there is no edge between i and j, we define $P_{ij} = 0$. Futhermore, we define the probability of staying at the poster i as $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$. Then as P defines a matrix, we can use this to find a random walk on the graph. In particular, this random walk will stay longer on the strong nodes (the ones that win a lot, as P_{ij} will be big for those nodes. So after 1000 iterations of the random walk, we can count how long the walk is on each node. This defines a ranking. I got the following ranking:
10-9-3-4-2-11-8-5-1-6-12-7

CONCORDET WINNER

The Concordetwinner is the poster that wins of every poster[1]. So I just calculated for every pair (i, j) how many times i wins of j and j wins of i. If i wins more often of j than j wins of i, then i is the Concordet winner. So I defined the funtion $a(i, j)$ counting how many times j wins of i, and then the difference $a(j, i) - a(i, j)$ for every pair i, j. It turned out that this value is nonnegative for every $j \neq i$ for $i = 10$. This means that 10 is the Concordet winner. Unfortunately, the Concordet method does not give a global ranking. Secondly, there is only a Concordetwinner if the value we calculated is actually positive for all j, but it still gives a good indication of the best poster.

BORDA COUNT

To use Borda count to find the optimal ranking, one gives score to every winner and loser. I chose to give a score of 1 if a poster wins and 0 if it loses. After assigning scores to every pairwise comparison, I add all the scores per poster and then divide by the number of times they are compared with another poster. Note that in the case of pairwise comparison, it does not matter which values for winning and losing I choose, because it is scale invariant. Furthermore, it is justified to divide by the number of times a poster is compared, as it will make up for the possibility that some posters could be voted more for (hence they can have a higher score).
Note that in the original application of Borda Count, evey vote consists of a total ranking for all posters. As every poster has been voted for the same amount of times, you do not have to divide by this number, as they are all equal. The total ranking we get is the following:
10-3-4-2-11-5-6-8-9-7-1-12. Hence poster 10 wins using Borda count.

KENNEMY OPTIMAZATION

Kennemy optimazation tries to find the optimal ranking such that the number of mismatches is minimal. Here the number of mismatches is defined as the number of times that some one votes that poster i is better than poster j, while in the ranking j is better than i for some $i, j \in \{1, \ldots, 12\}$. Kennemy optimazation is an NP-hard problem: it just tries every permutation of the twelve posters and calculates the number of mismatches. To approximate this number I used some kind of greedy method. I started with an arbitrary permutation, then for any two pairs (that may be overlapping) and switched the pairs and find out for which resulting permutation the number of mismatches is minimal. Then I repeated this process until the number of mismatches was stationary. I did this entire process 40 times, and the minimal number of mismatches I received was 15. The possible permutations for this are:

- 10 - 11 - 3 - 2 - 4 - 5 - 8 - 6 - 1 - 7 - 9 - 12
- 10 - 11 - 3 - 2 - 4 - 5 - 8 - 6 - 1 - 7 - 12 - 9
- 10 - 11 - 3 - 2 - 4 - 5 - 8 - 9 - 6 - 1 - 7 - 12
- 10 - 11 - 3 - 2 - 5 - 4 - 8 - 6 - 1 - 7 - 12 - 9
- 10 - 11 - 3 - 2 - 5 - 4 - 8 - 9 - 6 - 1 - 7 - 12
- 10 - 11 - 3 - 2 - 5 - 4 - 9 - 8 - 6 - 1 - 7 - 12
- 10 - 11 - 3 - 5 - 2 - 4 - 8 - 6 - 1 - 7 - 9 - 12
- 10 - 11 - 3 - 5 - 2 - 4 - 8 - 6 - 1 - 9 - 7 - 12
- 10 - 11 - 3 - 5 - 2 - 4 - 9 - 8 - 6 - 1 - 7 - 12

Although it is not sure that this is the actual optimal solution, it would make sense that the optimal solution has 15 mismatches, because every time we end up with almost the same permutation. Secondly, when calculating the number of mismatches for the permumations recieved by the other methods, one gets that the number of mismatches is bigger than 15. So we can be sure that we did not miss those ones in trying to find the minimal number of mismatches.
The fact that there are so many possibilities for the same number of mismatches is probably because the dataset is quite small. However, also when the dataset grows bigger, it is still possible to have this many permutations, as for example if two posters are almost as good as eachother, then both of them will win equally often of eachother, making it independent which one should be ranked better.

CONCLUSION

As we have seen, all methods give the same winner. So for this dataset it is safe to say that poster 10 is the best poster. However, from the second place on there is no definite second poster anymore. As the Concordet method only gives the winner, we cannot say much about that. However, the rankings of the random walk and borda count are more similar than if one compares any of these rankings with the ranking of kennemy optimazation (for so far we can be sure that we have found the optimal solution). So all together, apart from the winner, the rankings are very different, so one should choose the ranking according to one's wishes to find a good ranking.

REFERENCES

[1] Concordet Internet voting servise. Concordet completion in civs.

[2] Devavrat Shah Sahand negahban, Sewoong Oh. Rank centrality: Ranking from pair-wise comparisons. 2014.

[3] Sewoong Oh. Ranking from pairwise comparisons.

FUTURE RESEARCH

For future research one could apply these methods on other datasets and see if the conclusions still hold for those datasets. Furthermore, one could also see if one can find a theoretical reason for the results I found. Lastly, one can describe when every method can be used best, hence when every method gives the best result according to your preferences.