Analyzing Consensus Rankings in Large Datasets

Hilson Shrestha, Kartik Nautiyal

1 Introduction

Parallel Coordinates Plot (PCP) is a great choice when it comes to visualizing multi-variate numerical data and it can be used to visualize ranking scenarios like group-decision making where decision-makers build a consensus ranking that best represents the collection of base rankings.

Hindalong et. al [3] used the PCP technique to visualize the ranking problem since it helps users to visualize how individual candidate is ranked by different rankers. However, to visualize how a consensus is reached is not possible using just PCP because of a limitation: Comparison is limited to adjacent axes. To solve this problem, the system integrates box plots and strip plots for the analysis of consensus ranking. However, the system was designed for a limited number of rankers and candidates and does not scale up for a large dataset.

Another approach to visualizing consensus is to use color-coded stacked histograms [1]. While this technique allows the comparison of non-adjacent axes and could be effectively used to compare consensus ranking with all the base rankings, the system would not scale up for a large dataset with a large number of rankers. Users will have to scroll through all the base rankings to see how consensus alights with it.

To solve this, we attempt to integrate parallel coordinates plots with idioms like box plots or strip plots. We hope that our design will make it possible to analyze the degree of agreement of base rankings with consensus ranking. Furthermore, to visualize and inspect a large dataset, we attempt to exploit the spatial distortion technique [4].

2 One-sentence description

Combining parallel coordinates with other idioms like strip plot or box plot, we attempt to make analysis of consensus ranking in large dataset much easier.

3 Project Type

Development

4 Audience

Who is the audience for this project? How does it meet their needs? What happens if their needs remain unmet?

This project can be used in scenarios where consensus ranking is generated:

1. Ranking participants in competitions for final decision.
2. Ranking of colleges and universities
3. Ranking of students by admissions committee
4. Ranking of candidates for scholarships distribution

5 Approach

5.1 Details

What is your approach?

Our approach is to integrate PCP with idioms like strip plot or box plot. PCP will be used to visualize differences in different consensus rankings and box plot or strip plot will be used to analyze the agreement on each candidate by different base rankers. To visualize large dataset, we intend to extend this visualization by integrating fish eye view [5]. For detailed information on each candidate's ranking, we will integrate additional information view to get insight of how individual ranker ranked the candidate.

5.2 Evidence for Success

Why do you think it will work?

We think that the visualization will work because:

1. Box-plot or strip plot will help explain the distribution of individual ranker's perspective on each candidate.
2. Combined box-plot or strip plot on all candidates will give the general idea of how much the consensus ranking agrees to all the base rankings combined.
3. Implementation of fish eye view will help inspect individual candidate

6 Best-case Impact Statement

In the best-case scenario, what would be the impact statement (conclusion statement) for this project?

In the best case, our visualization will be able to explain degree of consensus in the consensus ranking. If there are multipe variations of consensus rankings, users will be able to differentiate between consensus rankings with more agreement consensus and rankings with less agreement.

7 Major Milestones

1. Generate a dataset with large number of rankers and candidates.
2. Generate multiple consensus rankings for comparison.
3. Implementation of visualization with PCP and box-plot or strip plot
4. Implementation of hover on each ranking to visualize base ranker's decision on each candidate.
5. Fish eye view for large dataset exploration

8 Obstacles

8.1 Major obstacles

1. Interactive visualization of agreement on each candidate
2. Generate dataset of rankers and candidates with atleast one consensus ranking

8.2 Minor obstacles

1. Fisheye view
2. Manipulate ranking by user

9 Resources Needed

What additional resources do you need to complete this project?

1. Code to generate large number of base rankings which have atleast some agreement 2. Code to generate different variations of consensus rankings 3. Test the clutter/understandability of box-plot vs strip plot

10 5 Related Publications

List 5 major publications that are most relevant to this project, and how they are related

1. Hindalong applied combination of different idioms to support group decision making for limited number of rankers and alternatives [3]
2. Sarkar proposes a metaphor of rubber sheet stretching for viewing large dataset in small space [4]
3. Tomininski implemented Fisheye view in tree graphs [5]
4. Bok extends parallel coordinates with color coded stacked histograms plot to overcome 2 major problems of PCP [1]
5. Gratzl ranks items based on attributes with different scales and includes PCP for comparison [2]

11 Define Success

What is the minimum amount of work necessary for this work be publishable?

If we can extend this project with user study to see if users can understand the degree of agreement in consensus ranking, then we believe we will have enough amount of work for publication.

References

[1] J. Bok, B. Kim, and J. Seo. Augmenting parallel coordinates plots with color-coded stacked histograms. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[2] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics*, 19(12):2277–2286, 2013.

[3] E. Hindalong, J. Johnson, G. Carenini, and T. Munzner. Towards rigorously designed preference visualizations for group decision making. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pages 181–190. IEEE, 2020.

[4] M. Sarkar, S. S. Snibbe, O. J. Tversky, and S. P. Reiss. Stretching the rubber sheet: A metaphor for viewing large layouts on small screens. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 81–91, 1993.

[5] C. Tominski, J. Abello, F. Van Ham, and H. Schumann. Fisheye tree views and lenses for graph visualization. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 17–24. IEEE, 2006.