Visualizations to support trade-off comparisons

Mehdi Chakhchoukh, Nadia Boukhelifa, Anastasia Bezerianos

English Abstract—In agronomy, domain experts use complex simulation models to understand real-world phenomena and plan biological studies. They use these models to generate big quantities of data that would help them make decisions while taking into account often conflicting socio-economical and environmental impacts. Such decision-making tasks can be very difficult for domain experts especially considering that they work with ranges and groups of points. They need to compare relatively big groups of points to weigh each solution space's pros and cons. This work aims to lay the foundation for generating design guidelines for visual group comparison in the context of trade-off analysis. As a first step towards this goal, we conducted a series of workshops to characterize the comparison needs in the trade-off analysis processes, particularly, in the case of group comparisons. Findings from these workshops confirm the need for group comparisons and show that our participants engaged in two high-level types of group comparisons: *composite* comparison and *ranking* comparison. We also found that our participants needed a modular granularity not only to manipulate and compare groups but also to compare point by point. Finally, we observed a need to aggregate the data and compare groups using quantitative metrics such as weighted means and variance.

1 Introduction

Comparison is a critical task for multi-criteria decision-making and, more specifically, for trade-off analysis. Trade-off analysis is a method of evaluating the advantages and disadvantages of different options or alternatives in a decision-making process. It is used to find a good compromise between conflicting goals. For example, in the case of wheat fertilization agronomists want to find a good balance between fertilizer doses, crop yield, and the impact on the environment. Experts are often comparing and ranking different alternatives, by taking into account the goals they want to optimize (e.g., a production strategy of higher fertilizer and yield, vs. one with lower yield but less environmental impact). These comparisons can become complex when experts need to keep track of many such objectives to optimize.

As we have found no clear definition for trade-off comparisons in the current data visualization and analytical tools literature, in the context of trade-off analysis we consider that : a comparison is a **relative assessment of elements grounded in a context comprised of the elements compared, the whole body of data, and the user preferences.** In other words, to compare trade-offs, one must consider their differences and similarities over the multidimensional space and their position in the dataset. This assessment is through the user preferences/goals, and it must also consider the tension between these different goals that may be in opposition.

- *Mehdi Chakhchoukh and Anastasia Bezerianos are with Univ. Paris-Saclay, CNRS, INRIA.*
  E-mail: first-name.last-name@universite-paris-saclay.fr.
- *Nadia Boukhelifa: is with UMR MIA-Paris, AgroParisTech, INRAE, Univ. Paris-Saclay.*
  E-mail: nadia.boukhelifa@inrae.fr.

Previous work in fields such as agronomy [2] [3] has shown that domain experts often work with *groups of points* when exploring large trade-off spaces. For example, wine experts may have two sets of points (or recipes) that have similar aromatic profiles but with different fermentation strategies (as a result of different fermentation times, initial nitrogen quantities, temperature profiles, etc).

Comparing these groups remains a challenging task, especially as the number of groups to be compared grows, the number of points within groups can also be large, and each point behaves differently in the multidimensional space. We hypothesise that visualization can help support trade-off comparisons by showing clearly the pros and cons of each group, and helping users articulate and refine their goals during the exploration.

Existing visualization tools in the literature for multi-criteria decision-making [6] [7] propose ranking systems to compare and assess the different items in a dataset and make informed decisions. For example, Lineup [6] uses a weighted scoring system to calculate the rank of a single item represented as a row and compare how the changes in the goal priorities can change the ranking. It is however difficult to compare groups of items with these tools,* as there is no way to group the items and consider them together as one entity while still having a clear visibility of the individual items constituting the groups.

Thus we conducted workshops with three groups of expert and non-expert users to better understand their needs in terms of comparisons and, more specifically, in terms of comparison of groups of points.

2 Related Work

In their work [5], [4] Gleicher et al. introduce a framework for understanding comparison tasks. They also introduce a taxonomy for comparisons which inclues three approaches: juxtaposition, superposition, and explicit encoding of relationships. They explain that one of the main challenges when working with comparisons is scalability, as comparison problems scale in both the complexity and quantity of the objects being compared

Most current multi-objective optimization decision-making tools rely on ranking for comparing solutions and picking the best ones. In their work [6] Gratzl et al. present a visual analytic tool called Lineup for ranking. Lineup uses weighted scores to calculate and display rankings in the form of a tabular representation. The users can update the weights and filter the data using ranges in real-time. In addition to comparing the solutions using rankings, the tool enables the comparison of rankings with a split view.

Although Lineup is good at ranking and at comparing individual items, it is not designed to compare groups of items that together constitute one solution and have complex relationships within the group (e.g., complementary, coverage of a range of values, group aggregation). It may also be difficult to ensure the coverage of data as the effect of changing weights is difficult to predict.

In their work, [7] Pajer et al. improved on Lineup by visualizing how the differences in weight influence the outcome of the ranking.

3 Work with domain experts

We revisited videos of three past trade-off analysis explorations conducted in previous work [2] [3] with domain experts, and looked for comparisons (6h of videos). We noticed that the experts mainly work with groups of items (where each item constitutes a stand-alone solution, e.g. a fertilisation strategy or a wine recipe), that they try to *rank*. In an attempt to cover as many types of possible comparisons during trade-off analysis tasks, we started considering everyday decision tasks. Apart from simple choice tasks, we identified cases where each data item is part of a *composite* solution and cannot be considered a stand-alone solution on its own. Examples include: creating a diet plan out of a dataset where each row is a food item, or putting together a sports team where each individual has different strengths and weaknesses.

The type of comparison depends on how one frames the problem and their data. Given a dataset of foods with their nutritional values, one may compare two groups of points that are rich in protein. For example, a vegan high-protein group of points and an animal produce high-protein group of points (*ranking comparison*). After they pick one of the two groups, one may merge that high-protein group with a group of points that represent a full diet and compare how the merging affects the nutritional balance of the diet. Their goal with the diet group may be to cover all the macro-nutrients a human needs while avoiding some micronutrients that would be undesirable or in opposition with one's goals *composite comparison*.

To better understand users' needs when conducting comparisons, we organized workshops (exploratory sessions and interviews) with three groups of users. Group 1 data contained points that were alternative solutions (and would thus more likely lead to ranking comparisons) and Group 2 and 3 had a dataset that contained points that were considered part of a solution (and could thus lead to composite comparisons).

Group 1: was comprised of two machine learning experts. They used a ML benchmark suit called Openml-cc18 [1]. The objective of this exploration is to examine and understand the potential trade-offs and connections among various attributes within a set of benchmark Machine Learning (ML) datasets. One specific trade-off revolves around the number of features present in a benchmark dataset, its inherent dimensionality (excluding noise), and the average correlation between features. As the number of features increases, it is expected that the intrinsic dimensionality will also increase. Conversely, the average correlation between features is likely to decrease.

We loaded their data in a visual data exploration tool called VisProm [3] and interviewed them about a past exploration and their needs in terms of comparisons, how they conduct them, and what they look for when performing these comparisons. This session lasted two hours.

Group 2: Was comprised of a participant with specific dietary needs. We loaded the USDA Foundation foods dataset (https://fdc.nal.usda.gov/download-datasets.html) in the visual data exploration tool [3] and then we asked them to perform a list of exploratory tasks with a think-aloud protocol:

Creating two diets (diets that have about 40 items and that have at least 50% dissimilarities): One that maximizes energy (calories) while minimizing sugar and fat.This diet should not have more than 5 carbohydrates-heavy foods (>5g). And another that Maximizes energy (calories) while minimizing sugar and carbohydrates. This diet should not have more than 5 fat-heavy foods (>5g). These are trade-off analysis and construction tasks that require comparisons and were inspired by the participant's self-reported dietary constraints.

Then the participant compared the 2 diets in terms of nutrients (iron, magnesium, fibers, vitamins,...).

Group 3: Was comprised of a participant with very restrictive dietary needs. We conducted two sessions. The first one was the same as the one conducted with group 2. In the second session the participant was asked to adapt the diet created during the 1st session to their personal dietary needs but maintain a similar nutritional balance (in this case the diet needed to become vegetarian and dairy free). This additional task is a more complex trade-off analysis task as it adds more conflicting objectives. It is also more personal as it is based on the participant's needs and thus it was added in the hopes of seeing more interesting, organic and explicit examples of comparisons being used during a trade-off analysis of composite solutions.

4 Workshops Results

4.1 Group 1: machine learning experts

The machine learning experts reported that they usually compare groups of points such as datasets and clusters. "...for this project, we were comparing data sets. What normally happens is that we try to compare groups inside the same data sets. So maybe cluster, maybe classes, stuff like that... Or methods like you may compare methods. In the last few works, we were mainly looking at specific trade-offs for the accuracy and the complexity of the models."

When asked about what they looked at during those comparisons they reported looking at the relative positions of the groups, "...Usually the relative position. You might be interested in assessing whether the two groups overlap for one metric or another for one dimension over another, or how much they overlap". When enquiring further about the precise metrics by which they compare groups such as ranges or specific values (minimum/maximum) they said that it is problem-dependent. They did, however, give some examples: in some cases, they may want to look at the relative bests in each group [this suggests the need for ranking the points in a group]; in other cases they compare the ranges and thresholds of the groups; finally they mentioned a case where they compared the means and variances of the groups.

4.2 Groups 2: diet creation I

The participant in that session worked with acceptable ranges and thus compared the groups of points generated from the ranges. Then they refined their selections little by little: They removed groups of points strategically by comparing subgroups of the same group and removing the ones deemed less useful. Then, they skimmed through the individual points and selected the ones to remove. To do that, they would compare the points within the same group with one another. These comparisons were conducted by looking at the difference between the values on each dimension.

4.3 Groups 3: diet creation II

The participant worked mainly with percentile ranges to identify foods that correspond to their criteria. This meant that they were constantly comparing the number of points in their selection vs the number of points outside the selection. Once they had a selection that was compliant with the requirements, they skimmed through the individual points to select the ones they liked the most. To do that they would compare points within one group with one another. These comparisons were conducted by looking at the difference between the values on each dimension. Even when they reached a good trade-off balance, they reported looking for diversity and complementary foods to constitute a diet that is pleasant to eat.

5 Conclusion

Comparison and ranking are important tasks in multi-criteria decision-making. Although a lot of work has been done for the ranking and comparison of individual points, ranking and comparison of groups of points remains an open challenge. Our observations from working with agronomy and ML experts have shown that they work primarily with solution spaces (i.e. groups of items as standalone or composite solutions). To better identify user needs for comparisons of groups, we conducted workshops with three groups of participants. These provided us with valuable insights into how trade-off comparisons are conducted, in what contexts, and on which objects. Such findings will inform future workshops with HCI and visualization experts, from which we will derive inspiration for creating and evaluating visual tools that specifically cater for group comparisons.

References

[1] B. Bischl, G. Casalicchio, M. Feurer, P. Gijsbers, F. Hutter, M. Lang, R. G. Mantovani, J. N. van Rijn, and J. Vanschoren. Openml benchmarking suites, 2021.

[2] N. Boukhelifa, A. Bezerianos, I. Cristian Trelea, N. Perrot, and E. Lutton. An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '19*, 2019.

[3] M. Chakhchoukh, N. Boukhelifa, and A. Bezerianos. Understanding How In-Visualization Provenance Can Support Trade-off Analysis. *IEEE Transactions on Visualization and Computer Graphics*, May 2022.

[4] M. Gleicher. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24:413–423, 2018.

[5] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10:289 – 309, 2011.

[6] 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 (InfoVis '13)*, 19(12):2277–2286, 2013.

[7] S. Pajer, M. Streit, T. Torsney-Weir, F. Spechtenhauser, T. Möller, and H. Piringer. Weightlifter: Visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):611–620, 2017.