DISTANCE-BASED AGGREGATION AND CONSENSUS FOR PREFERENCE-APPROVALS

Alessandro Albano [1], Mariangela Sciandra [1] and Antonella Plaia [1]

[1] Department of Economics, Business and Statistics, University of Palermo, (e-mail: alessandro.albano@unipa.it,mariangela.sciandra,antonella.plaia)

ABSTRACT: This paper proposes a distance-based aggregation and consensus method for preference-approvals, a type of preference data where individuals provide a list of approved alternatives in addition to a strict ranking. The proposed method aims to synthesize individual preference-approvals into a unified consensus representing the group's collective view. The consensus is the preference-approval, which minimizes the average distance with the whole set of voters. The proposed method has potential applications in group decision-making, recommendation systems, and social choice theory.

KEYWORDS: preference-approvals, preference aggregation, group decision-making, consensus

1 Introduction

In recent years, preference aggregation has received much attention due to its various applications in group decision-making, recommendation systems, and social choice theory. One type of preference data that has gained increasing popularity is preference-approvals, where individuals provide a list of approved alternatives in addition to a ranking (Brams & Sanver, 2009). In this paper, we propose a distance-based aggregation and consensus method for preference-approvals, which aims to synthesize individual preference-approval into a unified consensus representing the group's collective view. The proposed method finds the consensus as the preference-approval that minimizes the average distance with the whole set of voters. We employ a family of distances to evaluate the disagreement between preference-approvals and then use this to formulate an optimization problem to find the consensus preference-approval. This paper presents the notation and framework necessary to understand the proposed method describing the aggregation procedure. This method could advance preference aggregation and aid in practical decision-making scenarios.

2 Notation

Suppose a set of voters $V = \{v_1, \ldots, v_n\}$, with $n \geq 2$, are asked to order m different alternatives. The ranking π is a mapping function from the set of alternatives $X = \{x_1, \ldots, x_m\}$ to the set of ranks $\pi = \{P_\pi(x_1), \ldots, P_\pi(x_i), \ldots, P_\pi(x_m)\}$, where $P_\pi : X \longrightarrow \{1, \ldots, m\}$ assigns the rank of each alternative.

In the framework of preference-approval modelling, each preference ranking, π, is paired with an approval vector, A. For any given set X of alternatives, we define approvals by partitioning X into the set of approved alternatives G and the set of rejected alternatives U. We represent a voter's preference-approval profile by a top-down order of alternatives with a horizontal bar: alternatives above the bar are approved, and those below are rejected.

$$\begin{array}{c} x_3 \\ \hline x_2 \\ x_1 \\ x_4 \end{array}$$

The preference-approval above is codified as follows:

$$\pi_1 = (2,3,1,4) \quad A_1 = (0,0,1,0).$$

To evaluate the disagreement between preference-approvals, Erdamar *et al.* (2014) introduced a family of distances. Specifically, given a parameter $\lambda \in [0,1]$, they define a distance for preference-approvals, denoted by d_λ, as a mapping from pairs of preference-approval profiles to the interval $[0,1]$.

$$d_\lambda\big((\pi_1, A_1), (\pi_2, A_2)\big) = \lambda d_K(\pi_1, \pi_2) + (1-\lambda) d_H(A_1, A_2) \tag{1}$$

where (π_1, A_1) and (π_2, A_2) are two preference-approval profiles for the same set of alternatives X of size m, d_K and d_H are respectively the Kemeny and Hamming distance. In a recent study, Albano *et al.* (2022) presented a generalized version of d_λ, denoted as D^r_λ. This extended distance measure incorporates a power-weighted mean as an aggregation function and accounts for discordance between pairs of items in the preference-approval profiles.

3 Aggregation procedure

Given a $n \times 2m$ matrix Π, whose l-th row represents the preference-approval associated with the l-th judge, the consensus preference-approval $(\hat{\pi}, \hat{A})$ is found by minimizing the average distance function d_λ for fixed λ:

$$(\hat{\pi}, \hat{A})_\lambda = \underset{(\pi,A) \in P^m}{\arg\min} \sum_{l=i}^{n} d_\lambda((\pi^{(l)}, A^{(l)}), (\pi, A)), \tag{2}$$

where P^m is the universe of all preference-approvals with m objects.

By construction, the minimization of d_λ entails the simultaneous minimization of both rank and approval distances. Therefore, the problem is reduced to finding $\hat{\pi}$ and $\hat{A}$ such that:

$$(\hat{\pi} = \underset{\pi \in S^m}{\arg\min} \sum_{l=i}^{n} d_K(\pi^{(l)}, \pi),\ \hat{A} = \underset{A \in \{0,1\}^m}{\arg\min} \sum_{l=i}^{n} d_H(A^{(l)}, A)). \tag{3}$$

where S^m is the universe of the permutations (with ties) of m elements, and $d_H(A^{(l)}, A)$ and $d_K(\pi^{(l)}, \pi)$ are respectively the Hamming and the Kemeny distance between the preference and the approval part of the l-th row and the candidate consensus.

To find the Kemeny optimal ranking $\hat{\pi}$, we rely on the work of D'Ambrosio *et al.* (2015), who provided two accurate algorithms, called QUICK and FAST, for identifying the median ranking following the Kemeny approach. To find the approval consensus, $\hat{A}$, we compute the median approval vector by calculating the element-wise median of the binary approval matrices for all judges. In other words, we calculate the median of each column of the binary approval matrix, resulting in a final approval vector representing the consensus among the judges.

4 Case study

This section presents a case study, using data from the Eurobarometer*, website to demonstrate the effectiveness of the proposed method. The data consists of 27 rows (one per EU member country) and 9 columns representing alternatives concerning social values such as x_1: Equality between women and men, x_2: Fight against discrimination, x_3: Tolerance and respect for diversity, x_4: Solidarity among EU States, x_5: Solidarity between the EU and poor countries, x_6: Protection of human rights, x_7: Freedom of religion, x_8: Freedom of movement, and x_9: Freedom of speech. To obtain preference-approvals, alternatives are ranked in order of popularity for each country, and those that received more votes than the national average were considered acceptable. We used a hierarchical clustering procedure based on d_λ (with $\lambda = 0.75$) and found that the EU countries can be separated into two large clusters. Cluster 1 mainly comprises Western European countries (Austria, Belgium, Denmark, France, Italy, Luxembourg, Malta, Netherlands, Portugal, Spain, and Sweden). In contrast, Cluster 2 is manly composed of Eastern European countries (Bulgaria, Croatia, Cyprus, Czech Rep., Estonia, Finland, Germany, Greece, Hungary, Ireland, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia). The consensus procedure has been applied to aggregate preference-approvals within

*https://europa.eu/eurobarometer/surveys/detail/2612.

each cluster and facilitates the interpretation. The two consensus preference-approvals are:

Cluster 1	Cluster 2
$x_1\, x_9$	x_6
x_6	x_9
———	x_8
x_4	x_4
x_2	———
x_3	$x_1\, x_3$
x_8	x_2
x_5	x_5
x_7	x_7

The two consensus clusters show different levels of agreement on certain alternatives. For instance, Cluster 1 consensus shows a higher preference for equality between women and men. In contrast, Cluster 2 consensus shows a higher preference for the solidarity between EU Member States and freedom of movement. Overall, the two consensus preference-approvals provide a more detailed and nuanced picture of how the EU countries express their views on the nine alternatives proposed.

5 Conclusions

In conclusion, this paper proposes a distance-based approach for aggregating and reaching a consensus on preference-approvals, providing a solution for extracting a common preference from a group with diverse preferences. The approach offers a framework for achieving consensus among individuals with diverse preferences and can help improve decision-making processes' effectiveness and efficiency. Moreover, this algorithm could be used within preference learning algorithms to make predictions. In future work, we aim to extend this approach to the generalized distance function presented by Albano *et al.* (2022), thus providing an algorithmic solution to achieving consensus through the extended preference-approval distance.

References

ALBANO, A., GARCÍA-LAPRESTA, JOSÉ LUIS., PLAIA, A., & SCIANDRA, M. 2022. A family of distances between preference-approvals. *Annals of Operations Research*, 1–29.

BRAMS, STEVEN J., & SANVER, M. REMZI. 2009. *Voting Systems that Combine Approval and Preference*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 215–237.

D'AMBROSIO, ANTONIO, AMODIO, SONIA, & IORIO, CARMELA. 2015. Two algorithms for finding optimal solutions of the Kemeny rank aggregation problem for full rankings. *Electronic Journal of Applied Statistical Analysis*, **8**(2), 198–213.

ERDAMAR, BORA, GARCÍA-LAPRESTA, JOSÉ LUIS, PÉREZ-ROMÁN, DAVID, & SANVER, M REMZI. 2014. Measuring consensus in a preference-approval context. *Information Fusion*, **17**, 14–21.