A dimensionality Reduction Approach for Qualitative Preference Aggregation

Quentin Brabant, Miguel Couceiro, Fabien Labernia, Amedeo Napoli

A Dimensionality Reduction Approach for Qualitative Preference Aggregation

Quentin Brabant[1], Miguel Couceiro[1], Fabien Labernia[2], Amedeo Napoli[1]

[1] LORIA (CNRS, Inira Nancy Grand Est - Université de Lorraine
{quentin.brabant, miguel.couceiro, amedeo.napoli}@inria.fr
[2] LAMSADE (CNRS - Université Paris-Dauphine)
fabien.labernia@dauphine.fr

1 Qualitative preference aggregation models

In this paper we briefly present a method for reducing the dimensionality of data in a qualitative preference aggregation framework. For a more complete description of this approach, see [4]. For an alternative approach based on rough sets theory, see [1].

We consider the following setting. X is a set of alternatives that are evaluated according to a set of criteria represented by there indices: $[n] = \{1, \dots, n\}$. For an alternative $x \in X$ we denote by $(x_1, \dots, x_n) \in L^n$ the tuple of the evaluations of x in each crieterion. L is called the *evaluation space*, and is a distributive lattice for which we denote respectively by 0 and 1 the minimal and maximal element. We consider a binary preference relation $\preccurlyeq$ between the alternatives that can be expressed in terms of a *utility function*:

$$\forall x, y \in X : x \preccurlyeq y \Rightarrow U(x) \leq U(y),$$

where $U : X \to L$ associates a global evaluation on L to each alternative, and is obtained through the aggregation of the evaluations in criteria by a Sugeno integral $\mathcal{S}_\mu : L^n \to L$. In other words we have $U(x) = \mathcal{S}_\mu(x_1, \dots, x_n)$. The Sugeno integral defined over distributive lattices [3], is expressed

$$\mathcal{S}_\mu(x_1, \dots, x_n) = \bigvee_{I \subseteq [n]} \mu(I) \bigwedge_{i \in I} x_i,$$

where $\mu : 2^{[n]} \to L$ a capacity, that is to say a non-decreasing set function on $[n]$, with $\mu(\emptyset) = 0$ and $\mu([n]) = 1$. Capacities (and Sugeno integrals) are defined by a value on L for each subset of $[n]$, and therefore carry an intrinsic complexity, that grows exponentially with n. We now consider a set $\mathscr{D} = \{(\mathbf{x}^1, y^1), \dots, (\mathbf{x}^m, y^m)\} \subseteq L^n \times L$, where each $\mathbf{x}^i = (x_1^i, \dots, x_n^i) \in L^n$ is a tuple of evaluations in n critieria, and y^i is a utility value associated to $\mathbf{x}^j$. We want to learn a Sugeno integral $\mathcal{S}_\mu$ that generalizes these data. Ideally this function would be such that $\mathcal{S}_\mu(\mathbf{x}^j) = y^j$ for any $j \in \{1, \dots, m\}$. However, it is very common that no such function exists: in that case $\mathscr{D}$ is said to be *inconsistent*, and we aim at learning a Sugeno integral that realizes the prediction of y^j for each element, with an error as low as possible. Because of the nature of capacities, this optimization problem is on 2^n variables, and is therefore hard to solve when a high number of criteria is considered.

2 Dimensionnality reduction based on quality measure

By a *quality measure* over $\mathscr{D}$ we mean a degree with which $\mathscr{D}$ satisfies a certain hypothesis. In this presentation we consider two of such measures.

The first quality measure is the *monotonicity degree*, that is, the ratio of pairs $\{i, j\} \subseteq \{1, \ldots, m\}$ that satisfy the following condition

$$y^i > y^j \Rightarrow \exists k \in [n] : y_k^i > y_k^j.$$

This condition can be seen as a generalization of the Pareto condition to partially ordered evaluation spaces. The second quality measure is the *compatibility degree*, that is, the ratio of pairs satisfying the the condition

$$\exists \mathcal{S}_\mu : [\mathcal{S}_\mu(\mathbf{x}^i) = y^i \text{ and } \mathcal{S}_\mu(\mathbf{x}^j) = y^j]. \tag{1}$$

This condition is justified by results from [2] that apply only when L is totally ordered. Indeed it can be shown that $\mathscr{D}$ is consistent if and only if (2) is true for any pair from $\mathscr{D}$. Moreover, for a given pair this condition can be checked in a linear time w.r.t. n. Hence, provided that L is totally ordered, the compatibility degree is both theoretically meaningful and practically interesting. If L is not totally ordered, the monotonicity degree is the quality measure that makes sense.

The principle of the algorithm for dimensionality reduction that we propose is to iteratively remove a criterion, in order to minimize the decrease of the quality of the dataset at each step. Criteria are deleted until it is impossible to remove a criterion without decreasing the quality of the data below a certain ratio α. This algorithm was tested on empirical data [1] and allowed a reduction of the number of criteria from 7 to 3. Aggregation models trained on original data and on data reduced to 3 criteria showed to have similar accuracy. On the other hand, models trained on data with only 2 criteria left had significantly worse accuracy, suggesting that a reduction to 3 criteria constitutes the best compromise between simplicity and accuracy for these data.

Future research work should include further empirical studies and should aim to determining a procedure for deciding the optimal value of α, currently being set by hand.

References

1. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research, 138(2), 247–259 (2002)
2. Prade, H., Rico, A., Serrurier, M., Raufaste, E.: Elicitating Sugeno integrals: Methodology and a case study. Proceedings of the 10th European Conference Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 712–723 (2009)
3. Couceiro, M., Marichal, J.-L.: Characterizations of discrete Sugeno integrals as polynomial functions over distributive lattices. Fuzzy Sets and Systems 161(5), 694–707 (2010)
4. Brabant, Q., Couceiro, M., Labernia, F., Napoli, A.: Une approche de réduction de dimensionnalité pour l'agrégation de préférences qualitatives. Proceedings of the 16th Conférence francophone sur l'Extraction et la Gestion des Connaissances, 345–350 (2016)

[1] Tripadvisor: `http://sifaka.cs.uiuc.edu/~wang296/Data/index.html`