HAL
open science

Une approche de réduction de dimensionnalité pour l'agrégation de préférences qualitatives

Quentin Brabant, Miguel Couceiro, Fabien Labernia, Amedeo Napoli

To cite this version:

Quentin Brabant, Miguel Couceiro, Fabien Labernia, Amedeo Napoli. Une approche de réduction de dimensionnalité pour l'agrégation de préférences qualitatives. 16ème Journées Francophones Extraction et Gestion des Connaissances, EGC 2016, Jan 2016, Reims, France. pp.345-350. hal-01404913

HAL Id: hal-01404913

https://inria.hal.science/hal-01404913v1

Submitted on 29 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une approche de réduction de dimensionnalité pour l'agrégation de préférences qualitatives

Quentin Brabant *, Miguel Couceiro *, Fabien Labernia **, Amedeo Napoli *

* LORIA (CNRS - Inria Nancy Grand Est - Université de Lorraine)
Équipe Orpailleur, Bâtiment B, campus scientifique
B.P. 239, 54506 Vandoeuvre-les-Nancy, France
{quentin.brabant,miguel.couceiro,amedeo.napoli}@inria.fr
**LAMSADE (CNRS - Université Paris-Dauphine)
Place du Maréchal de Lattre de Tassigny
75016 Paris, France
fabien.labernia@dauphine.fr

Résumé. Nous présentons une méthode de réduction de dimensionnalité pour des données de préférences multicritères lorsque l'espace des évaluations est un treillis distributif borné. Cette méthode vise à réduire la complexité des procédures d'apprentissage d'un modèle d'agrégation sur des données qualitatives. Ainsi nous considérons comme modèle d'agrégation l'intégrale de Sugeno. L'apprentissage d'un tel modèle à partir de données empiriques est un problème d'optimisation à 2^n paramètres (où n est le nombre de critères considérés). La méthode de réduction que nous proposons s'appuie sur l'observation de certaines relations entre les éléments de ces données, et nous donnons des premiers résultats d'applications.

1 Introduction

Rappelons d'abord le cadre dans lequel s'applique la méthode de réduction de dimensionnalité que nous allons présenter.

Soit un ensemble de n critères que l'on représente par un ensemble d'entiers noté $[n] = \{1, \ldots, n\}$, et un ensemble d'alternatives noté X, dont chaque élément peut être évalué relativement à chaque critère, sur un espace d'évaluation que nous appellerons L. On définit $\varphi : X \to L^n$ comme la fonction associant à tout $x \in X$ ses évaluations par critère. On appelle *valeur d'utilité* d'une alternative la valeur donnée par une *fonction d'utilité* $U : X \to L$, qui évalue la désirabilité d'une l'alternative. On peut déterminer un ordre de préférence $\preccurlyeq$ sur X, par rapport à U :

$$\forall x, x' \in X : x \preccurlyeq x' \Leftrightarrow U(x) \leq U(x').$$

Autrement dit, une alternative est *préférée ou indifférente à* une autre si et seulement si elle possède une valeur d'utilité supérieure ou égale. Souvent, en aide multicritère à la décision, L est un intervalle numérique. Nous nous intéressons ici aux cas où L est un treillis distributif

borné. On suppose que la valeur d'utilité d'une alternative peut s'obtenir par agrégation de ses évaluations par critère, c'est à dire qu'il existe une fonction d'agrégation A telle que $U(x) = A(\varphi(x))$ pour tout $x \in X$. Cette fonction d'agrégation est appelée *modèle d'agrégation de préférences*.

Exemple 1. *On considère un ensemble de voitures X, évaluées selon trois critères, par exemple Marque, Couleur et Prix. Soit un espace d'évaluation partiellement ordonné $L = \{b, m, g, u\}$, dont les éléments correspondent resp. aux évaluations "mauvais" (b), "moyen" (m), "bon" (g) et "inconnu" (u), où $b < m < g$ et $b < u < g$, avec m et u incomparables. L est donc bien un treillis distributif borné. Soit x et x' deux voitures appartenant à l'ensemble X, pour lesquelle les évaluations par critère sont les suivantes : $\varphi(x) = (g, m, m)$, $\varphi(x') = (m, b, m)$. Les valeurs d'utilité de x et x' sont alors données par $U(x) = A(g, m, m)$, $U(x') = A(m, b, m)$, où A est une fonction d'agrégation, qui reste à déterminer.*

Ce que nous appellerons ici *élicitation* du modèle d'agrégation de préférences est la recherche de la fonction d'agrégation A représentant au mieux la manière dont un décideur évalue une alternative en fonction de son évaluation sur plusieurs critères. Les données dont on dispose pour réaliser une telle élicitation peuvent être formalisées comme un ensemble :

$$\mathscr{D} = \{(\mathbf{y}^1, y^1), \ldots, (\mathbf{y}^m, y^m)\} \subseteq L^n \times L.$$

Chaque couple $(\mathbf{y}^i, y^i)$ représente une alternative avec ses évaluations par critère $\mathbf{y}^i = (y_1^i, \ldots y_n^i) \in L^n$ et sa valeur d'utilité $y^i \in L$.

En général, le modèle d'agrégation est représenté par une intégrale floue ; une intégrale de Choquet lorsque L est un intervalle numérique, ou une intégrale de Sugeno, lorsque L est un treillis distributif borné. Nous nous intéressons à ce second cas. Dans la suite de cet article, L désignera donc toujours un treillis distributif borné, et les opérateurs $\wedge$ et $\vee$ noteront respectivement l'opération d'infimum et de supremum. Pour des références générales sur les intégrales floues, voir par exemple Beliakov et al. (2007); Bouyssou et al. (2013); Grabisch et al. (2009).

2 Agrégation de préférences avec l'intégrale de Sugeno

L'intégrale de Sugeno est une fonction d'agrégation qui a été introduite par Sugeno (1974) et qui se définit par rapport à une capacité. Nous appelons *capacité* sur un ensemble $[n]$ une fonction $\mu\colon 2^X \to L$ telle que

— $\mu(\emptyset) = 0$ et $\mu(X) = 1$, et

— pour tous $I, I' \subseteq X$ on a $I \subseteq I' \implies \mu(I) \leq \mu(I')$.

Ici, le rôle de la capacité est d'associer à chaque combinaison de critères une valeur assimilable à son importance dans le processus d'agrégation, X est donc l'ensemble des critères $[n]$. Une capacité permet de représenter les phénomènes de synergie et de redondance entre les critères ; c'est à dire, en terme de préférences, le fait qu'une bonne évaluation simultanée de deux attributs (ou plus) soit spécialement désirable, ou au contraire superflue.

Soit une capacité $\mu\colon 2^{[n]} \to L$. L'*intégrale de Sugeno* définie par rapport à μ, notée par $\mathcal{S}_\mu$, s'exprime sous la forme d'un polynôme latticiel (voir Couceiro et Marichal (2010)) :

$$\mathcal{S}_\mu(y_1, \ldots, y_n) = \bigvee_{I \subseteq [n]} \mu(I) \wedge \bigwedge_{i \in I} y_i,$$

2.1 Élicitation de l'intégrale de Sugeno

Les difficultées que présentent l'élicitation de l'intégrale de Sugeno dépendent principalement de deux paramètres, à savoir : le fait que L soit un ordre partiel ou total [1], et le fait que les données utilisées pour l'élicitation soient consistantes ou non.

On dit que $\mathscr{D}$ est *consistant* s'il existe $\mathcal{S}_\mu$ telle que $\forall i \in [m] : \mathcal{S}_\mu(\mathbf{y}^i) = y^i$. On dit que $(\mathbf{y}^i, y^i)$ est *compatible* avec $(\mathbf{y}^j, y^j)$ s'il existe $\mathcal{S}_\mu$ telle que $\mathcal{S}_\mu(\mathbf{y}^i) = y^i$ et $\mathcal{S}_\mu(\mathbf{y}^j) = y^j$. Prade et al. (2009) montrent que dans le cas où L est totalement ordonné, $\mathscr{D}$ est consistant si et seulement si tous les éléments de $\mathscr{D}$ sont compatibles deux à deux. De plus, on peut vérifier la compatibilité de deux éléments de $\mathscr{D}$ par une procédure en temps linéaire par rapport au nombre d'attributs. Il est alors possible de vérifier la consistence de $\mathscr{D}$ par une procédure de complexité quadratique par rapport à la taille de $\mathscr{D}$. Cela permet également de définir un degré de consistance de $\mathscr{D}$, valué sur $[0, 1]$, de la manière suivante :

$$consistance(\mathscr{D}) = \frac{|\{\{a, b\} \subseteq \mathscr{D} : a \text{ et } b \text{ compatibles}\}\}|}{|\{\{a, b\} \subseteq \mathscr{D}|}.$$

Dans cet article nous nous intéressons principalement aux données inconsistantes, que L soit totalement ordonné ou non. L'élicitation de $\mathcal{S}_\mu$ est alors un problème d'optimisation où la valeur à minimiser est une mesure de l'erreur de prédiction de $\mathcal{S}_\mu$ sur $\mathscr{D}$ (voir Section 2.1.1). Ce problème d'optimisation est complexe car il revient à déterminer la valeur de μ sur chaque sous-ensemble de critères, au nombre de 2^n, n étant le nombre de critères considérés. Nous ne connaissons aucune méthode permettant de résoudre ce problème d'optimisation en un temps raisonnable, c'est pourquoi nous avons eu recours au recuit simulé. Le recuit simulé est une méta-heuristique reposant sur le parcours stochastique de l'espace des solutions d'un problème cible. À chaque solution potentielle est associé un coût qui est la variable à minimiser. Ici chaque solution potentielle correspond à une capacité sur l'ensemble des critères, dont le coût est égal à l'erreur sur $\mathscr{D}$.

2.1.1 Mesure de l'erreur

Habituellement, on définit l'erreur d'une prédiction comme la distance entre une valeur prédite par un modèle et une valeur effective. Ici, le calcul de l'erreur pose problème ; puisque les valeurs appartiennent à un ensemble ordinal, on ne connaît pas à proprement parler la distance entre deux éléments de cet ensemble.

La première solution pour mesurer l'erreur consiste à considérer que la distance entre deux éléments $a, b \in L$ est égale à la longueur du plus court chemin entre ces deux éléments dans L, la distance entre deux éléments "voisins" étant arbitrairement considérée comme valant 1. Ainsi l'erreur de prédiction d'une intégrale $\mathcal{S}_\mu$ sur un couple $(\mathbf{y}, y)$ s'exprime par $d(\mathcal{S}_\mu(\mathbf{y}), y)$, et on peut définir l'erreur de prédiction de cette même intégrale sur l'ensemble $\mathscr{D}$ comme le cumul des erreurs de prédiction sur chaque élément de $\mathscr{D}$, normalisé par la taille de l'ensemble.

La seconde manière de mesurer l'erreur ne prend pas de notion de distance en compte et est présente dans la littérature sous le nom de *pairwise error*. L'ensemble $\mathscr{D}$ définit un ordre sur ses éléments par les valeurs d'utilité $y^1, \ldots, y^m$. Nous comparons cet ordre à l'ordre donné par $\mathcal{S}_\mu(\mathbf{y}^1), \ldots, \mathcal{S}\mu(\mathbf{y}^m)$, qui correspond à l'ensemble des prédictions du modèle. Afin de mesurer

1. un ordre total se représente par une chaîne, qui est un treillis distributif borné

la dissimilarité entre ces ordres, on considère chaque paire $\{(\mathbf{y}^i, y^i), (\mathbf{y}^j, y^j)\} \subseteq \mathscr{D}$; lorsque la relation entre $\mathcal{S}\mu(\mathbf{y}^i)$ et $\mathcal{S}\mu(\mathbf{y}^j)$ est différente de la relation entre y^i et y^j, on considère que le modèle a généré une erreur d'ordonnancement. La mesure de l'erreur peut alors être définie comme la proportion de paires mal ordonnées par le modèle.

Cette seconde méthode fournit sans doute le meilleur indice de la justesse d'un modèle par rapport $\mathscr{D}$; cependant elle demande d'effectuer un nombre de vérifications d'ordre $\mathcal{O}(m^2)$, c'est pourquoi lors de la procédure de recuit simulée, nous calculons le coût de chaque solution à l'aide de la première méthode.

3 La réduction de dimensionnalité

Description de la méthode

Ici, l'enjeu de la réduction de dimensionnalité est de supprimer, avant l'élicitation, une partie des critères considérés, tout en évitant les répercutions significativement négatives sur les modèles d'agrégations générés à partir des données réduites.

Notre méthode fonctionne par itération d'un processus de choix et de suppression d'un seul critère. Á chaque itération, les critères sont triés en fonction de l'impact qu'aurait leur suppression sur une mesure de qualité des données, et le dernier critère est supprimé, à moins qu'une certaine condition d'arrêt ne soit atteinte (décrite plus loin). Nous proposons deux mesures de la qualité des données.

La première est le degré de consistance des données. Il parait en effet raisonnable de supposer qu'un haut degré de consistance a un impact positif sur la précision du modèle entraîné sur ces données. La seconde mesure est le degré de monotonicité des données. De la même manière que le degré de consistance mesure la proportion de paires compatibles par rapport à l'intégrale de Sugeno, le degré de monotonicité mesure la proportion de paires $\{(\mathbf{y}^i, y^i), (\mathbf{y}^j, y^j)\}$ qui respectent la contrainte suivante :

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

Cette contrainte est plus faible que la propriété de compatibilité selon l'intégrale de Sugeno. Ces mesures de qualités peuvent donc être vues comme des mesures du degré auquel les données satisfont les hypothèses sur lesquelles reposent nos modèles.

La suppression d'un critère provoque nécessairement la baisse (ou la stagnation) de la mesure de qualité des données ; notre condition d'arrêt consiste à définir un seuil α, correspondant au ratio maximal de perte acceptable de qualité lors de la suppression d'un critère. Notons $\mathscr{D}$ un l'ensemble des données avant une itération quelconque et $\mathscr{D}_{-c}$ ce même ensemble dont on retire le critère c. On considère que la suppression de c est acceptable si

$$\text{QUALITÉ}(\mathscr{D}) \leq \alpha \times \text{QUALITÉ}(\mathscr{D}_{-c}),$$

où $\alpha > 1$. Dans le cas où aucun critère ne vérifie cette condition, le processus de réduction de dimensionnalité renvoie pour résultat $\mathscr{D}$.

3.1 Application à des données empiriques

Données

Les données que nous avons utilisées proviennent de reviews d'utilisateurs sur des hôtels [2], auxquelles étaient parfois associées des évaluations de ces hôtels. Ces évaluations sont effectuées sur 7 critères, sur une échelle de 1 à 5, ce que nous avons traité comme un ensemble totalement ordonné $L = \{1, 2, 3, 4, 5\}$, tel que $1 < 2 < 3 < 4 < 5$.

Après sélection aléatoire d'un échantillon $\mathscr{D}_{learn}$ de 250 éléments parmi l'ensemble des données, nous avons appliqué notre méthode de suppression des critères selon trois mesures de qualité différentes : le degré de consistance, le degré de monotonie et une mesure aléatoire. La figure 1 montre la précision moyenne des modèles générés sur les données résultantes de la méthode de réduction, en fonction du nombre de critères restants et de la méthode employée pour la sélection des critères. Nous n'avons pas appliqué la condition d'arrêt, afin d'observer l'évolution des données et des modèles qui en résultaient jusqu'à 2 critères.

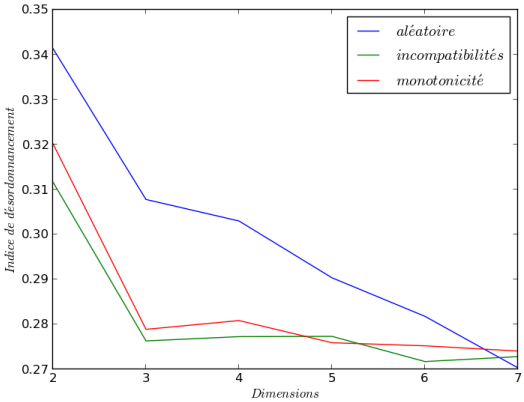

FIG. 1 – *Erreurs d'ordonnancement en fonction du nombre de critères pris en compte. Chaque point est le résultat d'une moyenne de 20 exécutions de la procédure.*

On constate que les deux variantes de la méthodes donnent de meilleurs résultats qu'une séléction aléatoire, ce qui indique que les mesures de qualités sont pertinentes et permettent toutes deux d'arriver au résultat souhaité. La précision des modèles d'agrégation ne semble en outre que peu impactée par la suppression de critères, jusqu'à un seuil de 3 critères.

Il semble donc qu'il existe un nombre optimal de critères à prendre en compte (du moins en ce qui concerne les données que nous avons traitées).

L'application de cette méthode à nos données produit de bons résultats : pour une valeur [3] de $\alpha = 1.5$, la procédure (dont les résultats ont été lissés grâce à une moyenne effectuée sur 20 exécutions du programme) supprime en moyenne 3.9 critères sur les 7, et les modèles générés donnent un indice de désordonnancement moyen de $0,28$, très proche de l'indice idéalement attendu (voir figure 1).

2. Tripadvisor Dataset : `http://sifaka.cs.uiuc.edu/~wang296/Data/index.html`.
3. α est ici un paramètre fixé par l'utilisateur.

4 Conclusion

Les méthodes de réduction de dimensionnalité que nous avons proposées ont donné des résultats prometteurs et bien que ceux-ci soient à relativiser, car partiellement dépendants des données et de l'heuristique utilisées, il apparaît que la sélection de critères est une approche prometteuse pour la réduction de la complexité dans le cadre de l'agrégation de préférences qualitatives.

Faute de données suffisantes, nous n'avons pas pu appliquer nos méthodes à différents contextes. En particulier, nous n'avions à notre disposition aucunes données de préférences multicritères dont l'espace d'évaluation était partiellement ordonné, contexte pour lequel nos modèles sont d'abord conçus, et des tests sur de telles données sont à envisager. D'une façon plus globale, nous espérons à l'avenir être en mesure de généraliser les résultats obtenus dans cet article. S'il parait raisonnable de supposer que la mesure des incompatibilités peut permettre de déduire l'importance des attributs au sein de tout ensemble de données, il est tout à fait possible que la condition d'arrêt que nous avons utilisée soit dépendante des données que nous avons considérées.

Références

Beliakov, G., A. Pradera, et T. Calvo (2007). *Aggregation functions : A guide for practitioners*, Volume 221. Springer.

Bouyssou, D., D. Dubois, H. Prade, et M. Pirlot (2013). *Decision Making Process : Concepts and Methods*. John Wiley & Sons.

Couceiro, M. et J.-L. Marichal (2010). Characterizations of discrete Sugeno integrals as polynomial functions over distributive lattices. *Fuzzy Sets and Systems 161*(5), 694–707.

Grabisch, M., J.-L. Marichal, R. Mesiar, et E. Pap (2009). *Aggregation functions*, Volume 127. Cambridge University Press.

Prade, H., A. Rico, M. Serrurier, et E. Raufaste (2009). Elicitating Sugeno integrals : Methodology and a case study. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 712–723. Springer.

Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. Tokyo Institute Tech.

Summary

We present a method for dimension reduction in multicriteria preference modeling, when taking a bounded distributive lattice as evaluation space. This method aims to reducing the complexity of learning an aggregating model on qualitative data. In this setting, the aggregation model of choice is the Sugeno integral. Learning such a model with qualitative data can thus be turned into an optimization problem with 2^n parameters (where n is the number of considered criteria). The reducing method that we propose is based on such qualitative data, and we will give some preliminary application results.