Robust Ordinal Regression for Collaborative Preference Learning with Opinion Synergies

Extended Abstract

Hugo Gilbert
Université Paris-Dauphine, Université PSL, CNRS, LAMSADE
75016, Paris, France
hugo.gilbert@lamsade.dauphine.fr

Mohamed Ouaguenouni
Sorbonne Université, CNRS, LIP6
75005, Paris, France
mohamed.ouaguenouni@lip6.fr

Meltem Öztürk
Université Paris-Dauphine, Université PSL, CNRS, LAMSADE
75016 Paris, France
meltem.ozturk@lamsade.dauphine.fr

Olivier Spanjaard
Sorbonne Université, CNRS, LIP6
75005, Paris, France
olivier.spanjaard@lip6.fr

ABSTRACT

This work focuses on a robust learning methodology in a collaborative filtering context. We wish to predict preferences between alternatives characterized by binary attributes, where each attribute represents the opinion of a reference user on the alternative. The model whose parameters we learn is general enough to be compatible with any strict weak order on the attribute vectors, thanks to the consideration of opinion synergies. Moreover, we accept not to predict some preferences if the data collected are not compatible with a reliable prediction. A predicted preference will be considered reliable if all the simplest models explaining the training data agree on it. Following the robust ordinal regression methodology, our predictions are based on an ordinal dominance relation between alternatives introduced by Fishburn and LaVille [11] which relies on an uncertainty set encompassing the possible values of the parameters of the multi-attribute utility function.

KEYWORDS

Collaborative Preference Learning; Robust Ordinal Regression; Multi-Attribute Utility Theory

ACM Reference Format:

Hugo Gilbert, Mohamed Ouaguenouni, Meltem Öztürk, and Olivier Spanjaard. 2023. Robust Ordinal Regression for Collaborative Preference Learning with Opinion Synergies: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 CF MEETS ROR

Collaborative Filtering (CF) aims at producing recommendations on items (movies, products, books, etc.) to a user by relying on the matrix of ratings provided by the users on subsets of items, without exogenous information about users or items [15]. CF methods are often partitioned into two categories: the neighborhood methods, which make recommendations based on similarity measures computed from the rating matrix, and the latent-factor models that build a representation of users and items in a common factor space to explain the ratings and make recommendations.

Differently, we characterize each item (also called alternative) by a vector of positive or negative opinions (also called attributes) of reference users on it. We denote by $\mathcal{U}$ the set of reference users, by $\mathcal{A}$ the set of alternatives, and by $\mathcal{F} = \{a_1, a_2, \ldots, a_n\}$ the set of binary attributes. Each attribute a_i corresponds to a reference user U_i and a_i equals 1 (resp. 0) if U_i has a positive (resp. negative) opinion on it. An alternative A is characterized by its attribute vector in $\{0, 1\}^n$ and can be viewed, by abuse of notation, as a subset of attributes. To predict the preferences of a (non-reference) user, we learn a Multi-Attribute Utility (MAU) function f [6, 8, 9, 14] defined on $\mathcal{A}$, with $f(A) > f(B)$ iff A is strictly preferred to B.

Following Fishburn and Lavalle [11], we consider a function f where the value $f(S)$ of a set S of attributes is an additive combination of parameters, one per subset A of S: $f(S) = \sum_{A \subseteq S} v_A$. Similarly to the closery related Choquet integral [1, 4, 16], this model is general enough to model positive or negative synergies between attributes [13]. Moreover, it makes it possible to model any strict weak ordering on the subsets of attributes. Yet, it is intractable as there is a combinatorial set of parameters v_A. Similarly to the k-additive variant of this model which only considers v_A for $|A| \leq k$, we only consider a restricted family $\theta \subseteq 2^{\mathcal{F}}$ of subsets A to keep a tractable set. Formally, $f(S) = \sum_{A \in \theta} I_A(S) v_A$, where $I_A(S) = 1$ if $S \subseteq A$ and 0 otherwise. We call this model the *θ-additive model* and we may also use the notation $f_{\theta,v}(A)$ instead of $f(A)$.

Example 1. *Let $\mathcal{F} = \{a_1, a_2, a_3, a_4\}$, $\mathcal{A} = \{0, 1\}^4$ and the user's preferences be the strict weak order $\succ$ given by:* $(0, 1, 1, 1) \succ (1, 0, 1, 1) \succ (1, 1, 0, 1) \succ (0, 0, 1, 1) \succ (0, 1, 0, 1) \succ (0, 1, 1, 0) \succ (1, 0, 0, 1) \succ (1, 0, 1, 0) \succ (1, 1, 0, 0) \succ (0, 0, 0, 1) \succ (0, 0, 1, 0) \succ (0, 1, 0, 0) \succ (1, 0, 0, 0) \succ (\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}) \sim (0, 0, 0, 0) \succ (\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{0})$ *with $A \sim B$ iff $\neg(A \succ B)$ and $\neg(B \succ A)$. These preferences can be explained by a negative synergy when $a_1 = a_2 = a_3 = 1$ (bold vectors). Interestingly, instead of using a 3-additive model with 14 parameters, one can use the θ-additive model with $\theta = \{\{a_1\}, \{a_2\}, \{a_3\}, \{a_4\}, \{a_1, a_2, a_3\}\}$ and $v_{\{a_1\}} = 1$, $v_{\{a_2\}} = 2$, $v_{\{a_3\}} = 3$, $v_{\{a_4\}} = 4$, $v_{\{a_1,a_2,a_3\}} = -10$.*

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

We use the framework of Robust Ordinal Regression (ROR) [see e.g., 5, 7] to learn *both* the set θ and the parameters v_A themselves (for $A \in \theta$). Based on a set $R = \{(A,B) \in \mathcal{A}^2 | A \succ B\}$ of strict pairwise preferences inferred from the user's ratings, the idea is to consider the polyhedron of parameter values of f compatible with R. In our θ-additive model, this polyhedron is the set V_θ^R defined by:

$$V_\theta^R = \{v : \theta \to \mathbb{R} \mid \forall (A,B) \in R,\ f_{\theta,v}(A) > f_{\theta,v}(B)\}.$$

A preference between A and B is then predicted if $f(A) > f(B)$ for all possible parameters values. We refer to this dominance relation as the θ-ordinal dominance relation.

Definition 1. For a given $\theta \in \Theta_R$, the θ-ordinal dominance relation, denoted by $\succ_\theta^R$, is defined by:

$$\forall A, B \in \mathcal{A},\ A \succ_\theta^R B \Leftrightarrow \forall v \in V_\theta^R,\ f_{\theta,v}(A) > f_{\theta,v}(B).$$

This dominance relation can be tested by linear programming by working on the polyhedron corresponding to V_θ^R.

To avoid an arbitrary choice of the parameter set θ, we consider an even more general dominance relation. We denote by Θ_R the set $\{\theta \mid V_\theta^R \neq \emptyset\}$, i.e., the θ's such that the preferences in R are consistent with a θ-additive function. In our setting, a preference is predicted if $f(A) > f(B)$ for all the simplest sets $\theta \in \Theta_R$, and for all possible values v_A for $A \in \theta$. Indeed, following the philosophical principle of parsimony that the simpler of two explanations is to be preferred (Occam's razor [3]), we focus on "simple" sets.

Simplest models are defined by relying on three criteria: First, we consider subsets θ that minimize the complexity of synergies between the attributes; through the *degree* of θ, i.e., $\max\{|S| : S \in \theta\}$. Second, if two different θ have the same *degree*, we prefer the one which minimizes the *size* $|\theta|$ [18]. Third, if two different θ have the same degree and size, we choose the one which minimizes $\sum_{S \in \theta} |S|$.

These criteria define a lexicographic binary relation on Θ_R, denoted by $\sqsubseteq_{lex}$. We call the $\theta \in \Theta_R$ which are minimal according to $\sqsubseteq_{lex}$, the *simplest* θ of R, and we denote by $\Theta_R^{\min}$ their set. Based on $\Theta_R^{\min}$, we extend the ordinal dominance relation to define the notion of *robust ordinal dominance relation*.

Definition 2. The robust ordinal dominance relation, denoted by $\succ_\Theta^R$, is defined as follows:

$$A \succ_\Theta^R B \iff \forall \theta \in \Theta_R^{\min},\ A \succ_\theta^R B.$$

Before going into details of the computational aspects of our method, let us mention two studies, close to ours: Domshlak and Joachims [8] consider a MAU function that is a sum of 4^n subutilities over subsets of attribute values and develop an efficient SVM approach ; Bigot et al. [2] study the use of generalised additively independent decompositions of utility functions [10, 12] and give a polynomial PAC-learner. Even if we have common points in the use of utility functions taking into account synergies between attributes, these two works do not focus on ordinal robustness.

2 COMPUTING THE ROBUST ORDINAL DOMINANCE RELATION

To compute the robust ordinal dominance relation, one could compute all the elements in $\Theta_R^{\min}$. Unfortunately, as the size of $\Theta_R^{\min}$ can be exponentially large, this methodology reveals computationally cumbersome. We show in this section how to bypass this difficulty by using the following proposition:

Proposition 1. *Given a set R of strict pairwise comparisons, and $\theta \in \Theta_R$, if $R' \subseteq R$, then we have: (i) $\theta \in \Theta_{R'}$; (ii) $A \succ_\theta^{R'} B \Rightarrow A \succ_\theta^R B$; (iii) $A \succ_\theta^R B \Rightarrow \neg(B \succ_\theta^{R'} A)$.*

Thanks to this result, we know that, for any pair X, Y of alternatives, $\Theta_{R\cup\{(X,Y)\}}$ and $\Theta_{R\cup\{(Y,X)\}}$ are subsets of Θ_R. Hence, to evaluate if a robust ordinal dominance relation holds between X and Y, one could proceed by examining the following conditions:

$$\text{(i) } \Theta_R^{\min} \cap \Theta_{R\cup\{(X,Y)\}} \neq \emptyset, \quad \text{(ii) } \Theta_R^{\min} \cap \Theta_{R\cup\{(Y,X)\}} \neq \emptyset.$$

If (i) holds but (ii) does not, we know that $X \succ_\Theta^R Y$ as all the $\theta \in \Theta_R^{\min}$ can only account for the fact that X is preferred to Y and not the reverse. If both (i) and (ii) hold, then neither $X \succ_\Theta^R Y$ nor $Y \succ_\Theta^R X$.

Proposition 2. *Given any pair X, Y of alternatives,*
$X \succ_\Theta^R Y$ iff $\Theta_R^{\min} \cap \Theta_{R\cup\{(X,Y)\}} \neq \emptyset$ and $\Theta_R^{\min} \cap \Theta_{R\cup\{(Y,X)\}} = \emptyset$.

To determine if $\Theta_R^{\min} \cap \Theta_{R\cup\{(X,Y)\}} = \emptyset$, we rely on the concept of unifying model:

Definition 3. The *unifying model* is defined as $\theta_R^* = \bigcup_{\theta \in \Theta_R^{\min}} \theta$.

Using this concept, we proceed using the following steps. We start by computing a superset $\widetilde{\theta}_R^*$ of θ_R^* such that $\max\{|S| : S \in \widetilde{\theta}_R^*\} = \max\{|S| : S \in \theta_R^*\}$ and an element $\hat{\theta} \in \Theta_R^{\min}$. This is done by using a procedure that combines different linear and integer programs. Then, we test whether $\Theta_R^{\min} \cap \Theta_{R\cup\{(X,Y)\}} = \emptyset$ by determining the feasibility of the set of constraints 1–4 below. This set of constraints is feasible iff an element of $\Theta_R^{\min}$ makes it possible to represent that $X \succ Y$. Note that $\widetilde{\theta}_R^*$ is useful to reduce the size of this program.

$$\sum_{S \in \widetilde{\theta}_R^*} (I_A(S) - I_B(S)) v_S \geq 1, \quad \forall (A,B) \in R \cup \{(X,Y)\} \tag{1}$$

$$\sum_{S \in \widetilde{\theta}_R^*} b_S \leq |\hat{\theta}| \tag{2}$$

$$\sum_{S \in \widetilde{\theta}_R^*} b_S |S| \leq \sum_{S \in \hat{\theta}} |S| \tag{3}$$

$$-b_S M \leq v_S \leq b_S M \text{ and } b_S \in \{0,1\} \quad \forall S \in \widetilde{\theta}_R^* \tag{4}$$

where $M = (2|\widetilde{\theta}_R^*| + |R| + 1) \times (|R| + 1)^{2|R|+3}$ is set such that if the values v_S can be set to satisfy all the preferences in $R \cup \{(X,Y)\}$, then there exist such values in $[-M, M]$ (see [17]). Every instantiation of variables (v_S, b_S) (for $S \in \widetilde{\theta}_R^*$) that satisfies the constraints corresponds to an element $\theta = \{S \in \widetilde{\theta}_R^* \mid b_S = 1\} \in \Theta_R \cup \{(X,Y)\}$. Interestingly, as $\max\{|S| : S \in \widetilde{\theta}_R^*\} = \max\{|S| : S \in \theta_R^*\}$ we are sure that any such element has minimum degree. Then, the constraints 2 and 3 ensure that such an element θ also belongs to $\Theta_R^{\min}$.

3 CONCLUSION

Our contribution is twofold. First, we have proposed a collaborative preference learning method that only uses ordinal data, often more reliable than numerical one. Second, we have developed a multiattribute preference elicitation procedure that is able to learn both a sparse set of parameters and their values (which ensures explainability). Numerical tests have been carried out on synthetic and real data to evaluate the richness and reliability of the predictions made, showing that our robust regression ordinal methodology provides an interesting compromise between accuracy and prediction rate.

ACKNOWLEDGEMENTS

We acknowledge a financial support from the project THEMIS ANR20-CE23-0018 of the French National Research Agency (ANR).

REFERENCES

[1] Silvia Angilella, Salvatore Corrente, and Salvatore Greco. 2015. Stochastic multi-objective acceptability analysis for the Choquet integral preference model and the scale construction problem. *European Journal of Operational Research* 240, 1 (Jan. 2015), 172–182. https://doi.org/10.1016/j.ejor.2014.06.031

[2] Damien Bigot, Hélène Fargier, Jérôme Mengin, and Bruno Zanuttini. 2012. Using and Learning GAI-Decompositions for Representing Ordinal Rankings. In *ECAI'2012 workshop on Preference Learning (PL 2012)*. Johannes Fürnkranz and Eyke Hüllermeier, 5–10.

[3] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. 1987. Occam's razor. *Information processing letters* 24, 6 (1987), 377–380.

[4] Roman Bresson, Johanne Cohen, Eyke Hüllermeier, Christophe Labreuche, and Michèle Sebag. 2020. Learning 2-additive Hierarchical Choquet Integrals with non-monotonic utilities. In *DA2PL 2020*.

[5] Johnnie R Charnetski and Richard M Soland. 1978. Multiple-attribute decision making with partial information: the comparative hypervolume criterion. *Naval Research Logistics Quarterly* 25, 2 (1978), 279–288.

[6] Wei Chu and Zoubin Ghahramani. 2005. Preference learning with Gaussian processes. In *Machine Learning, Proceedings of the Twenty-Second International Conference (ICML 2005), Bonn, Germany, August 7-11, 2005 (ACM International Conference Proceeding Series, Vol. 119)*, Luc De Raedt and Stefan Wrobel (Eds.). ACM, Bonn, Germany, 137–144.

[7] Salvatore Corrente, Salvatore Greco, Miłosz Kadziński, and Roman Słowiński. 2013. Robust ordinal regression in preference learning and ranking. *Machine Learning* 93, 2 (2013), 381–422.

[8] Carmel Domshlak and Thorsten Joachims. 2005. Unstructuring user preferences: efficient non-parametric utility revelation. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence.* AUAI Press, Edinburgh, Scotland, 169–177.

[9] Paolo Dragone, Stefano Teso, and Andrea Passerini. 2017. Constructive Preference Elicitation over Hybrid Combinatorial Spaces. https://doi.org/10.48550/ARXIV.1711.07875

[10] Peter C Fishburn. 1970. *Utility theory for decision making.* Technical Report. Research analysis corp McLean VA.

[11] P. C. Fishburn and I. H. Lavalle. 1996. Binary interactions and subset choice. *European Journal of Operational Research* 92 (1996), 182–192.

[12] Christophe Gonzales and Patrice Perny. 2005. GAI Networks for Decision Making under Certainty. In *19th International Joint Conference on Artificial Intelligence – workshop on advances in preference handling*. Edinburgh, United Kingdom, 100–105. https://hal.archives-ouvertes.fr/hal-01492604

[13] Michel Grabisch, Ivan Kojadinovic, and Patrick Meyer. 2008. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package. *European journal of operational research* 186, 2 (2008), 766–785.

[14] Ralph L Keeney, Howard Raiffa, and Richard F Meyer. 1993. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge university press, Cambridge, United Kingdom.

[15] Yehuda Koren, Steffen Rendle, and Robert Bell. 2022. *Advances in collaborative filtering.* Springer, 91–142.

[16] Jean-Luc Marichal and Marc Roubens. 2000. Determination of weights of interacting criteria from a reference set. *European journal of operational Research* 124, 3 (2000), 641–650.

[17] Christos H Papadimitriou. 1981. On the complexity of integer programming. *Journal of the ACM (JACM)* 28, 4 (1981), 765–768.

[18] Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. 2015. A survey of sparse representation: algorithms and applications. *IEEE access* 3 (2015), 490–530.