# Lattice-Theoretic Approach to Version Spaces in Qualitative Decision Making

Miguel Couceiro[1(✉)] and Tamás Waldhauser[2]

[1] LORIA (CNRS - Inria Nancy Grand Est - Université de Lorraine), Equipe Orpailleur – Bâtiment B, Campus Scientifique, B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France
miguel.couceiro@inria.fr

[2] Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, Szeged H-6720, Hungary
twaldha@math.u-szeged.hu

**Abstract.** We present a lattice-theoretic approach to version spaces in multicriteria preference learning and discuss some complexity aspects. In particular, we show that the description of version spaces in the preference model based on the Sugeno integral is an NP-hard problem, even for simple instances.

## 1 Motivation

We consider an instance of supervised learning, where given a set $\mathbf{X}$ of objects (or alternatives), a set $L$ of labels (or evaluations) and a finite set $S \subseteq \mathbf{X} \times L$ of labeled objects, the goal is to predict labels of new objects.

In the current paper, our motivation is found in the field of preference modeling and learning (prediction). We take the decomposable model to represent preferences over a set $\mathbf{X} = X_1 \times \cdots \times X_n$ of alternatives (e.g., houses to buy) described by $n$ attributes $X_i$ (e.g., price, size, location, color). In this setting, preference relations $\preceq$ are represented by mappings $U \colon \mathbf{X} \to L$ valued in a scale $L$, and called "overall utility functions", using the following rule:

$$\mathbf{x} \preceq \mathbf{y} \quad \text{if and only if} \quad U(\mathbf{x}) \leq U(\mathbf{y}).$$

In other words, an alternative $\mathbf{x}$ is less preferable than an an alternative $\mathbf{y}$ if the score of $\mathbf{x}$ is less than that of $\mathbf{y}$. This representation of preference relations is usually refined by taking into account "local preferences " $\preceq_i$ on each $X_i$, modeled by mappings $\varphi_i \colon X_i \to L$ called "local utility functions", which are then merged through an aggregation function $p \colon L^n \to L$ into an overall utility function $U$:

$$U(\mathbf{x}) = p\left(\phi(\mathbf{x})\right) = p\big(\varphi_1(x_1), \ldots, \varphi_n(x_n)\big). \tag{1}$$

T. Waldhauser—This research is supported by the Hungarian National Foundation for Scientific Research under grant no. K104251 and by the János Bolyai Research Scholarship.


A. Gammerman et al. (Eds.): SLDS 2015, LNAI 9047, pp. 234–238, 2015.
DOI: 10.1007/978-3-319-17091-6_18

Loosely speaking, $p$ merges the local preferences in order to obtain a global preference on the set of alternatives. In the qualitative setting, the aggregation function of choice is the *Sugeno integral* that can be regarded [10] as an *idempotent lattice polynomial function*, and the resulting global utility function (1) is then called a *pseudo-polynomial function*. This observation brings the concept of Sugeno integral to domains more general than scales (linearly ordered sets) such as distributive lattices and Boolean algebras. Apart from the theoretic interest, such generalization is both natural and useful as it allows incomparability amongst alternatives, a situation that is most common in real-life applications.

In this setting, we consider the learning scenario where given a training set consisting of pairs of alternatives together with their evaluations, we would like to determine all models (1) that are consistent with it. Formally we consider the following *supervised learning task*:

- **Given:**
  - a distributive lattice $L$ (evaluation space)
  - a training set $T$ of pairs $(\mathbf{x}, l_{\mathbf{x}}) \in \mathbf{X} \times L$
- **Find:** all pseudo-polynomials $U\colon \mathbf{X} \to L$ s.t. $U(\mathbf{x}) = l_{\mathbf{x}}$ for all $(\mathbf{x}, l_{\mathbf{x}}) \in T$.

In other words, we would like to describe the *version space* (see, e.g., [1,11]) in this qualitative setting, which asks for all lattice polynomial functions $p\colon L^n \to L$ and all local utility functions $\varphi_i\colon X_i \to L$, $i \in [n] := \{1, \ldots, n\}$, such that for every $(\mathbf{x}, l_{\mathbf{x}}) \in T$, $\mathbf{x} = (x_1, \ldots, x_n)$, we have $p\,(\phi(\mathbf{x})) = l_{\mathbf{x}}$.

*Remark 1.* This task of describing the version space was formalized mathematically in [3,5] as an interpolation problem. In this sense, we assume that the function $f_T$ determined by $T$ (i.e., $f_T(\mathbf{x}) = l_{\mathbf{x}}$ for each $(\mathbf{x}, l_{\mathbf{x}}) \in T$) is well-defined.

*Remark 2.* We would like to stress the fact that, despite motivated by a problem rooted in preference learning (see [7] for general background and a thorough treatment of the topic), our setting differs from the standard setting in machine learning. This is mainly due to the fact that we aim to describing utility-based preference models that are consistent with existing data (version spaces) rather than aiming to learning utility-based models by optimization (minimizing loss measures and coefficients) such as in, e.g., the probabilistic approach of [2] or the approach based on the Choquet integral of [12], and that naturally accounts for errors and inconsistencies in the learning data. Another difference is that, in the latter, data is supposed to be given in the form of feature vectors (thus assuming that local utilities over attributes are known a priori), an assumption that removes the additional difficulty that we face, namely, that of describing local utility functions that enable models based on the Sugeno integral that are consistent with existing data. It is also worth noting that we do not assume any structure on attributes and that we allow incomparabilities in evaluation spaces, which thus subsume preferences that are not necessarily rankings.

The complete description of the version space in the multicriteria setting still eludes us, but using some lattice-theoretic techniques developed in [3] (recalled in Section 2), we present in Section 3 explicit descriptions of version spaces when the local utility functions $\varphi_i\colon X_i \to L$ are known a priori.

## 2 When the Local Utility Functions are Identity Maps

Recall that a *polynomial function* over a distributive lattice $L$ is a mapping $p\colon L^n \to L$ that can expressed as a combination of the lattice operations $\wedge$ and $\vee$, projections and constants. In the case when $L$ is bounded, i.e., with a least and a greatest element denoted by 0 and 1, resp., Goodstein [8] showed that polynomial functions $p\colon L^n \to L$ coincide exactly with those lattice functions that are representable in *disjunctive normal form* (DNF for short) by

$$p(\mathbf{x}) = \bigvee_{I\subseteq[n]} \Big(c_I \wedge \bigwedge_{i\in I} x_i\Big), \text{ where } \mathbf{x} = (x_1, \ldots, x_n) \in L^n. \tag{2}$$

In fact he provided a canonical DNF where each $c_I$ is of the form $p(\mathbf{1}_I)$, where $\mathbf{1}_I$ denotes the "indicator function" of $I \subseteq [n]$.

Recall that the Birkhoff-Priestley representation theorem [6] states that we can embed $L$ into a Boolean algebra $B$. For the sake of canonicity, we assume that $L$ generates $B$, so that $B$ is uniquely determined up to isomorphism. We also denote the boundary elements of $B$ by 0 and 1. The complement of an element $a \in B$ is denoted by $a'$.

In [3], we showed that the version space when the local utility functions are identity maps (i.e., $X_i = L$ and the models are lattice polynomial functions) can be thought of as an interval lower and upper bounded by polynomials $p^-$ and $p^+$, resp., defined as follows. Let $T$ be a training set as above and consider the corresponding function $f_T\colon D \to L$ ($D$ comprises the first components of couples in $T$). Define the following two elements in $B$ for each $I \subseteq [n]$:

$$c_I^- := \bigvee_{\mathbf{a}\in D} \Big(f_T(\mathbf{a}) \wedge \bigwedge_{i\notin I} a_i'\Big) \qquad \text{and} \qquad c_I^+ := \bigwedge_{\mathbf{a}\in D} \Big(f_T(\mathbf{a}) \vee \bigvee_{i\in I} a_i'\Big).$$

and let $p^-$ and $p^+$ be the polynomial functions over $B$ given by:

$$p^-(\mathbf{x}) := \bigvee_{I\subseteq[n]} \Big(c_I^- \wedge \bigwedge_{i\in I} x_i\Big) \qquad \text{and} \qquad p^+(\mathbf{x}) := \bigvee_{I\subseteq[n]} \Big(c_I^+ \wedge \bigwedge_{i\in I} x_i\Big).$$

**Theorem 1 ([3]).** *Let $p\colon B^n \to B$ be a polynomial function over $B$ given by (2) and let $D$ be the set of the first components of couples in a training set $T$. Then $p|_D = f_T$ if and only if $c_I^- \leq c_I \leq c_I^+$ (or, equivalently, $p^- \leq p \leq p^+$).*

*Remark 3.* From Theorem 1 it follows that a necessary and sufficient condition for the existence of a polynomial function $p\colon B^n \to B$ such that $p|_D = f_T$ is $c_I^- \leq c_I^+$, for every $I \subseteq [n]$. Moreover, if for every $I \subseteq [n]$, there is $c_I \in L$ such that $c_I^- \leq c_I \leq c_I^+$, then and only then the version space over $L$ is not empty.

## 3 When the Local Utility Functions are Known A Priori

We now assume that the local utility functions $\varphi_i \colon X_i \to L$ are known a priori. So given a training set $T \subseteq \mathbf{X} \times L$, our goal to find all polynomial functions $p$ over $L$ such that the pseudo-polynomial function $U$ given by (1) is consistent with $T$, i.e., $U|_D = f_T$. Let us consider an arbitrary polynomial function $p$ over $B$ in its disjunctive normal form (2). The corresponding pseudo-polynomial function $U = p\left(\varphi_1, \ldots, \varphi_n\right)$ verifies $U|_D = f_T$ if and only if $p|_{D'} = f_{T'}$ where $D'$ is the set of first components of couples in $T' = \{(\phi(\mathbf{x}), f_T(\mathbf{x})) \colon (\mathbf{x}, f_T(\mathbf{x})) \in T\}$. Using the construction of Section 2 for $T'$, we define coefficients $c^-_{I,\phi}$ and $c^+_{I,\phi}$ for every $I \subseteq [n]$ as follows:

$$c^-_{I,\phi} := \bigvee_{\mathbf{a} \in D} \big(f_T(\mathbf{a}) \wedge \bigwedge_{i \notin I} \varphi_i(a_i)'\big) \quad \text{and} \quad c^+_{I,\phi} := \bigwedge_{\mathbf{a} \in D} \big(f_T(\mathbf{a}) \vee \bigvee_{i \in I} \varphi_i(a_i)'\big).$$

Denoting the corresponding polynomial functions by $p^-_\phi$ and $p^+_\phi$, Theorem 1 yields the following explicit description of version spaces.

**Theorem 2.** *Let $T \subseteq \mathbf{X} \times L$ and $D$ as before. For any $\varphi_i \colon X_i \to L\,(i \in [n])$ and any polynomial function $p \colon B^n \to B$ given by (2), we have that $U = p\,(\phi)$ verifies $U|_D = f_T$ if and only if $c^-_{I,\phi} \leq c_I \leq c^+_{I,\phi}$ for all $I \subseteq [n]$ (or, equivalently, $p^-_\phi \leq p \leq p^+_\phi$).*

*Remark 4.* Note that if there exist couples $(\mathbf{a}, f_T(\mathbf{a})), (\mathbf{b}, f_T(\mathbf{b})) \in T$ such that $f_T\left(\mathbf{a}\right) \neq f_T\left(\mathbf{b}\right)$ but $\phi(\mathbf{a}) = \phi(\mathbf{b})$, then the corresponding version space in the current multicriteria setting is void. We invite the reader to verify that this situation cannot occur if the condition of Theorem 2 is satisfied.

## 4 Concluding Remarks and Further Directions

Theorem 2 describes the version space for given local utility functions $\varphi_i \colon X_i \to L$. It still remains an open problem to determine all such local utility functions for which an interpolating polynomial function exists. Looking for interpolating polynomials over $B$, this amounts to solving the system of inequalities $c^-_{I,\phi} \leq c^+_{I,\phi}$for the unknown values $\varphi_i\left(a_i\right)$. As the following example illustrates, this is a computationally hard problem, even in the simpler case of *quasi-polynomial functions*, where it is assumed that $X_1 = \cdots = X_n = X$ and $\varphi_1 = \cdots = \varphi_n = \varphi$.

*Example 1.* We define an instance of our interpolation problem for any finite simple graph $G = (V, E)$. Let $L = \{0, 1\}$ be the two-element chain, let $X = V \dot\cup \{s, t\}$, let $D = \{(u, v) : uv \in E\} \cup \{(t, s), (s, s), (t, t), (s, t)\}$, and let $f_T\left(t, s\right) = 1, f_T\left(s, s\right) = 0, f_T\left(t, t\right) = f_T\left(s, t\right) = 1$ and $f_T\left(u, v\right) = 1$ for all $uv \in E$. Since there are only 4 polynomials over $L$ (namely 0, 1, $x_1 \wedge x_2$ and $x_1 \vee x_2$), it is easy to verify that $p\left(\varphi\left(x_1\right), \varphi\left(x_2\right)\right)$ interpolates $f_T$ if and only if $p = x_1 \vee x_2$, $\varphi\left(s\right) = 0, \varphi\left(t\right) = 1$ and $\{v \in V : \varphi\left(v\right) = 1\}$ is a covering set of the graph $G$. Hence, already in this simple case, it is an NP-hard problem to describe the whole version space, as it involves finding all covering sets (in part., all minimal covering sets) in a graph.

Despite the above example, the corresponding decision problem (is there an interpolating pseudo-polynomial?) might be tractable. Also, in some special cases, such as when all the sets $X_i$ as well as $L$ are finite chains and one considers only order-preserving local utility functions $\varphi_i \colon X_i \to L$ (which is the case in many applications), it might be feasible to construct the full version space. These problems constitute topics of further research of the authors.

## References

1. Cornuéjols, A., Miclet, L.: Apprentissage artificiel - Concepts et algorithmes. Eyrolles (2010)
2. Cao-Van, K., De Baets, B., Lievens, S.: A probabilistic framework for the design of instance-based supervised ranking algorithms in an ordinal setting. Annals Operations Research **163**, 115–142 (2008)
3. Couceiro, M., Dubois, D., Prade, H., Rico, A., Waldhauser, T.: General interpolation by polynomial functions of distributive lattices. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012. CCIS, vol. 299, pp. 347–355. Springer, Heidelberg (2012)
4. Couceiro, M., Waldhauser, T.: Axiomatizations and factorizations of Sugeno utility functions. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **19**(4), 635–658 (2011)
5. Couceiro, M., Waldhauser, T.: Interpolation by polynomial functions of distributive lattices: a generalization of a theorem of R. L. Goodstein. Algebra Universalis **69**(3), 287–299 (2013)
6. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order. Cambridge University Press, New York (2002)
7. Fürnkranz, J., Hüllermeier, E. (eds.): Preference learning. Springer, Berlin (2011)
8. Goodstein, R. L.: The Solution of Equations in a Lattice. Proc. Roy. Soc. Edinburgh Section A 67, 231–242 (1965/1967)
9. Grätzer, G.: General Lattice Theory. Birkhäuser Verlag, Berlin (2003)
10. Marichal, J.-L.: Weighted lattice polynomials. Discrete Mathematics **309**(4), 814–820 (2009)
11. Mitchell, T.: Machine Learning. McGraw Hill (1997)
12. Tehrani, A.F., Cheng, W., Hüllermeier, E.: Preference Learning Using the Choquet Integral: The Case of Multipartite Ranking. IEEE Transactions on Fuzzy Systems **20**(6), 1102–1113 (2012)