



Version Space Learning for Possibilistic Hypotheses

Henri Prade, Mathieu Serrurier

► To cite this version:

Henri Prade, Mathieu Serrurier. Version Space Learning for Possibilistic Hypotheses. 17th European Conference on Artificial Intelligence (ECAI 2006), European Coordinating Committee for Artificial Intelligence (ECCAI); Italian Association of Artificial Intelligence, Aug 2006, Riva del Garda, Italy. pp.801-802. hal-03361589

HAL Id: hal-03361589

<https://hal.science/hal-03361589>

Submitted on 1 Oct 2021

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.



Distributed under a Creative Commons Attribution 4.0 International License

Version space learning for possibilistic hypotheses

Henri Prade and Mathieu Serrurier¹

Abstract. In this paper, we are interested in learning stratified hypotheses from examples and counter-examples associated with weights that express their prototypical importance. It leads to an extension of the well-known version space learning framework. In order to do that, we emphasize that the treatment of positive and negative examples in version space learning is reminding of a bipolar revision process recently studied in the setting of possibilistic information representation. Bipolarity appears when the positive and negative sides of information are specified in a distinct way. Then, we use the possibilistic bipolar representation setting, which distinguishes between what is guaranteed to be possible, and what is simply not impossible, as a basis for extending version space learning to examples associated with possibility degrees. It allows us to define a formal framework for learning layered hypotheses.

1 Introduction

Version space learning [3] is a general framework for concept learning. It proposes to identify all the hypotheses coherent with a set of examples or counter examples. The way for treating positive and negative examples during the learning process reminds a bipolar revision process, that is at work when positive information is said apart from negative information [1]. Negative information states what is known to be impossible, what is rejected, while positive information states what is possible for sure, feasible or satisfactory. Examples and counter examples w. r. t. an if-then rule are instances of positive and negative information. When a bipolar view of information is maintained, consistency between positive and negative information, namely what is positively assessed is included in what is not negatively assessed, should be enforced when new information is received. We first point out the similarity between version space learning and bipolar revision, when learning from a set of examples and counter-examples. Then, examples are associated with possibility degrees that describe how much a positive example is guaranteed to be possible or how much a negative example is not impossible. By taking advantage of bipolar possibilistic revision, we propose an original algorithm for version space learning in order to identify the set of layered hypotheses that are sound w. r. to t. examples.

2 Version space and bipolar information

Let \mathcal{X} denote a *feature space*. In the bipolar [1] possibilistic setting, two $[0, 1]$ -valued possibility distributions over \mathcal{X} are considered : δ describes to what extent configurations are guaranteed to be possible and π describes what is not impossible. These two distributions are consistent in the bipolar framework iff $\forall x \in \mathcal{X}, \delta(x) \leq \pi(x)$. Given δ and π , which respectively describe what is guaranteed

to be possible and what is not impossible, when a new piece of information of the form π_{new} or δ_{new} is presented, a revision process takes place : $\forall x \in \mathcal{X} \pi_{revised}(x) = \max(\pi(x), \delta_{new}(x))$ and $\delta_{revised}(x) = \min(\pi_{new}(x), \delta(x))$. Let us describe the version space learning framework from the bipolar point of view. Let $\mathcal{U} = \{0, 1\}$ be a *concept space*, where 1 means that the example of the concept is positive (guaranteed possible i.e. $\delta(x) = 1$), and 0 means that the example is negative (totally impossible i.e. $\pi(x) = 0$). An hypothesis h is a mapping from \mathcal{X} to \mathcal{U} . \mathcal{H} denotes the hypotheses space. The version space framework takes as input a set $\mathcal{S} = \{(x_i, u_i)_{i=1, \dots, m} \text{ s.t. } x_i \in \mathcal{X} \text{ and } u_i \in \mathcal{U}\}$ of m training examples, and a hypothesis space \mathcal{H} . The set \mathcal{H} is supposed to be equipped with a partial preorder \succeq expressing generality, formally: $h_1, h_2 \in \mathcal{H}$. $h_1 \succeq h_2$, iff $\{x \in \mathcal{X} | h_1(x) = 1\} \supseteq \{x \in \mathcal{X} | h_2(x) = 1\}$. An hypothesis h is *sound* w. r. t. a training example $(x, u) \in \mathcal{S}$ iff $h(x) = u$. If $\forall (x_i, u_i) \in \mathcal{S}$, h is sound w.r.t (x_i, u_i) , then h is said to be sound w. r. t. \mathcal{S} . This framework defines the *version space* $\mathcal{V} = \{h \in \mathcal{H} | h \text{ is sound with } \mathcal{S}\}$. Version space learning aims at identifying the upper and the lower bounds of this version space \mathcal{V} . The upper bound \mathcal{V}_g will contain the most general hypotheses, i.e., the ones that classify more examples, whereas the lower bound \mathcal{V}_s will contain the most specific ones, i.e., the hypotheses that classify less examples, both being compatible with the training examples. The general algorithm for learning version space consists in updating \mathcal{V}_s and \mathcal{V}_g step by step by considering an increasing set of examples. If the example is positive, all the hypotheses of \mathcal{V}_g that are not coherent are removed, since they cannot be generalized, and all the hypotheses of \mathcal{V}_s that do not cover the examples are minimally generalized in order to cover it. On the contrary, if the example is negative the hypotheses that are incoherent with the example are removed if they are in \mathcal{V}_s or minimally specialized if they are in \mathcal{V}_g .

We can view a hypothesis as a possibility distribution, which states what feature configurations x are possible : $\forall h \in \mathcal{H}, \forall x \in \mathcal{X}, \mu_h(x) = h(x)$. Then, given a set of hypotheses, we define two possibility distributions : the most specific one $\forall H \in 2^{\mathcal{H}}, \forall x \in \mathcal{X}, \delta_H(x) = \min\{\mu_h(x); h \in H\}$ and the most general one $\forall H \in 2^{\mathcal{H}}, \forall x \in \mathcal{X}, \pi_H(x) = \max\{\mu_h(x); h \in H\}$. Let $\pi_{\mathcal{S}}$ and $\delta_{\mathcal{S}}$ be the possibility distributions that correspond respectively to the most general and the most specific distribution revised by the examples in \mathcal{S} . The set of hypotheses in \mathcal{V}_s contains the most specific hypotheses that cover all the positive examples and no negative examples. It means that these hypotheses must identify, if possible, only the situations that are *guaranteed to be possible* w. r. t. the set of examples and then $\delta_{\mathcal{S}} \leq \delta_{\mathcal{V}_s}$. In the same way, since \mathcal{V}_g contains the most general hypotheses that cover all the positive examples and no negative ones, these hypotheses describe the situations that are *not impossible* w. r. t. the set of examples and then $\pi_{\mathcal{S}} \leq \pi_{\mathcal{V}_g}$.

¹ IRIT - CNRS, 118 route de Narbonne, 31062 Toulouse Cedex 09, France
email: prade,serrurier@irit.fr

3 Weighted extension of version space learning

Alg. 1 SpecializeStratifiedHypothesis

Require: $h_w = \{(h_1, \alpha_1), \dots, (h_n, \alpha_n)\}$ with $\alpha_1 \geq \dots \geq \alpha_n$

Require: $\langle x, 0, \alpha \rangle$

- 1: Let i such that $(h_i, \alpha_i) \in h_w$ $\alpha_i \geq \alpha$ and $\nexists j$ such that $\alpha_i > \alpha_j \geq \alpha$
 - 2: **for all** (h_j, α_j) such as $j > i$ **do**
 - 3: **if** $h_j(x) = 1$ **then**
 - 4: turn h_j in (h_j, α_j) into one of its minimal specialization
 - 5: **if** $h_i(x) = 1$ **then**
 - 6: turn (h_i, α_i) into (h'_i, α_i) where h'_i one of the minimal specialization of h_j
 - 7: **if** $\alpha_i \neq \alpha$ **then**
 - 8: add (h_i, α) in h_w
-

Alg. 2 GeneralizeStratifiedHypothesis

Require: $h_w = \{(h_1, \alpha_1), \dots, (h_n, \alpha_n)\}$ with $\alpha_1 \geq \dots \geq \alpha_n$

Require: $\langle x, 0, \alpha \rangle$

- 1: **if** $\exists i$ such that $(h_i, \alpha_i) \in h_w$ and $\alpha_i = \alpha$ **then**
 - 2: return all h_w in which h_i is turned into one of its minimal generalization
 - 3: **else**
 - 4: Let i such that $(h_i, \alpha_i) \in h_w$ $\alpha_i > \alpha$ and $\nexists j$ such that $\alpha_i > \alpha_j > \alpha$
 - 5: add (h_i, α_i) in h_w such that h'_i is a minimal generalization of h_i
-

We are now interested in version space learning with examples associated with weights corresponding to possibility degrees in a linearly ordered scale, here $[0, 1]$ for simplicity. These possibility values have a different meaning according as the examples are positive or negative. If the example is positive, the weight corresponds to a guaranteed possibility degree. If the example is negative, the weight refers to the possibility distribution π that describes what is not impossible. The examples are then described by $\mathcal{S}_w = \{(x, u, \alpha)\}$ with $\alpha = \delta(x)$ if $u = 1$ and $\alpha = \pi(x)$ otherwise. The hypothesis we want to learn describes a possibility distribution that associates a possibility degree to each example, which corresponds to the degree of compatibility of this example w. r. t. the target concept. The hypothesis is then a mapping from \mathcal{X} to $[0, 1]$. Given an hypothesis space \mathcal{H} , we now consider a stratified set of classical hypotheses $h_w = (h_1, \alpha_1), \dots, (h_n, \alpha_n)$ with $\forall i, h_i \in \mathcal{H}$ and $\alpha_1 \geq \dots \geq \alpha_n$. The result of h_w for x is $\forall x \in \mathcal{X}, h_w(x) = \max_i \{\min(h_i(x), \alpha_i)\}$ and the possibility distribution induced by h_w is $\forall x \in \mathcal{X}, \mu_{h_w} = h_w(x)$. The generality ordering on stratified hypotheses corresponds to the partial pre-order on possibility distributions, and thus $h_w^1 \leq h_w^2$ iff $\mu_{h_w^1} \leq \mu_{h_w^2}$.

A stratified hypothesis is sound w. r. t. an example if it does not underestimate the possibility of a positive example and does not overestimate the possibility degree of a negative example. So, h_w is sound w. r. t. \mathcal{S}_w iff $\forall (x, u, \alpha) \in \mathcal{S}_w = h_w(x) \geq \alpha$ if $u = 1$ and $h_w(x) \leq \alpha$ if $u = 0$. A totally positive (guaranteed possible) example corresponds to $(x, 1, 1)$, and a totally negative (impossible) example to $(x, 0, 0)$. On the contrary, the examples of the form $(x, 1, 0)$ and $(x, 0, 1)$ have no influence on the learning algorithm according to the definition of soundness. Indeed, an example of the

form $(x, 1, 0)$ means just that x is not guaranteed to be a positive example and $(x, 0, 1)$ means that we are certain that it is possible for x to be a positive example. The algorithm for learning the version space in the case of bipolar continuous examples follows the same structure as the classical version space learning algorithm by using specialization and generalization operators described respectively in Alg. 1 and Alg. 2.

Proposition 1 We note $\delta_{\mathcal{S}_w}$ and $\pi_{\mathcal{S}_w}$ the possibility distributions obtained by revising the bipolar distribution ($\forall x \in \mathcal{X}, \delta_{\mathcal{S}_w}(x) = 0, \forall x \in \mathcal{X}, \pi_{\mathcal{S}_w}(x) = 1$) with the new information of the form $\forall (x, u, \alpha) \in \mathcal{S} \pi_{new}(x) = \alpha$ if $u = 0$ or $\delta_{new}(x) = \alpha$ if $x = 1$. Given $\mathcal{V}_w = \langle \mathcal{V}_{sw}, \mathcal{V}_{gw} \rangle$ the continuous version space from the set of example \mathcal{S} , we have $\forall x \in \mathcal{X}, \delta_{\mathcal{S}_w}(x) \leq \delta_{\mathcal{V}_{sw}}(x)$ and $\pi_{\mathcal{S}_w}(x) \geq \pi_{\mathcal{V}_{gw}}(x)$.

This proposition shows that we can use version space learning for describing bounds of possibility distributions according to a set of descriptions of what is not impossible or guaranteed to be possible. Since it is described in intention in the version space case, the bounds may be too restrictive w. r. t. the maximal bounds described by bipolar revision due to the limited description capabilities of the hypothesis language.

4 Related works and concluding remarks

In this paper, we have emphasized the similarities that exist between the version space learning framework and the bipolar binary revision process. According to this remark, we have proposed an extension of version space learning that deals with examples associated with possibility degrees. These degrees correspond to non impossibility degrees for negative examples and to guaranteed possibility degrees for positive examples. What is learned is the bounding of the set of stratified hypotheses that are coherent with the data. As for the classical setting, it is obvious that bipolar version space learning cannot be used directly in applications, but can be used as a formal framework for possibilistic concept learning.

This approach could then be applied to possibilistic inductive logic learning [4], which learns stratified hypotheses from weighted examples. In this context, it suggests to modify the treatment of examples in order to consider weights as bipolar information and then to adapt the specialization and generalization operators as described in this paper. This approach is clearly different from the fuzzy version space in [2]. In particular, our approach makes a difference for the meaning, and then for the treatment, of the levels associated to the examples according to their positive or negative nature, which is not the case in fuzzy version space. Indeed, in the fuzzy case, a version space is learnt for each *alpha*-cut of the training set. What is learnt is then a stratified sets of version spaces and the soundness of hypotheses w. r. t. examples is classical, while what is learnt here is a version space of stratified hypotheses.

REFERENCES

- [1] D. Dubois, H. Prade, and P. Smets, 'Not impossible vs. guaranteed possible in fusion and revision', in *Proc. of the 6th. European Conference (ESQARU 2001)*, pp. 522–531. Springer-Verlag, (2001).
- [2] E. Hüllermeier, 'Inducing fuzzy concepts through extended version space learning', in *Proc. of 10th Int. Fuzzy Systems Association (IFSA-03) - LNAI 2715*, pp. 677–684, (2003).
- [3] T. Mitchell, 'Generalization as search', *Artificial Intelligence*, **18**, 203–226, (1982).
- [4] M. Serrurier and H. Prade, 'Possibilistic inductive logic programming', in *Proc. ECSQARU'05 - LNAI 3571*, pp. 675–686, (2005).