



HAL
open science

Comments on “Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization” by Eyke Hüllermeier

S Destercke

► **To cite this version:**

S Destercke. Comments on “Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization” by Eyke Hüllermeier. *International Journal of Approximate Reasoning*, 2014, 55, pp.1588 - 1590. 10.1016/j.ijar.2014.04.014 . hal-01076725

HAL Id: hal-01076725

<https://hal.science/hal-01076725>

Submitted on 22 Oct 2014

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.

Comments on “Learning from Imprecise and Fuzzy Observations: Data Disambiguation through Generalized Loss Minimization” by Eyke Hüllermeier

S. Destercke

CNRS, UMR 7253 Heudiasyc, Centre de recherche de Royallieu, 60205 COMPIEGNE

Abstract

Eyke Hüllermeier provides a very convincing approach to learn from fuzzy data, both about the model and about the data themselves. In the process, he links the shape of fuzzy sets with classical loss functions, therefore providing strong theoretical links between fuzzy modeling and more classical machine learning approaches. This short note discusses various aspects of his proposal as well as possible extensions. I will first discuss the opportunity to consider more general uncertainty representations, before considering various alternatives to the proposed learning procedure. Finally, I will briefly discuss the differences I perceive about a loss-based and a likelihood-based approach.

1. Introduction

The paper [3] by Eyke Hüllermeier is a crystal clear exposition of what I think are two important contributions: first, he lays bare the idea of data disambiguation, i.e., the fact that in case of data uncertainty learning about the model and about the data should go “hand in hand”; second he makes a formal connection between learning with fuzzy data and learning with specific loss functions. This latter contribution may even be more important than the first, as it provides fuzzy sets and fuzzy modeling with a clear interpretation in terms of loss functions, therefore bringing them closer to classical machine learning techniques.

The paper provides a generic method and illustrates it on simple examples, showing why it makes sense to use such a method. It opens wide area of potential research, in terms of problems to which the method can be applied to. Yet I think the view of the author as well as the idea of data disambiguation could be expanded in various ways, which I briefly discuss here. My background being mainly in uncertainty treatment, my comments will focus on that aspect rather than on aspects related to learning issues. I will address the following points:

Email address: sebastien.destercke@hds.utc.fr (S. Destercke)

- The extension of the proposed method to more general uncertainty models, with short examples;
- The possibility of using other "choice" rules, including non-precise ones;
- The disambiguation process and its ambiguity in the fuzzy case;
- The differences between a likelihood-based approach and a loss-based approach.

2. The use of more general uncertainty models

In the paper, data uncertainty is considered to be described by fuzzy sets, reduced to the case of classical sets or intervals in the first part to facilitate exposure. As fuzzy sets are known to be somehow of limited expressiveness (e.g., they do not include probabilities as particular cases), I think there may be some interest in extending the current proposal to more generic models of uncertainty. In particular, since the value $\mathbb{L}(Y, \hat{y})$ of the fuzzy loss function can be interpreted as a lower expectation (provided the possibility measure induced by the fuzzy set is read as an upper probability [2]), its extension to the case where the uncertainty of an observation Y_i is described by a (convex) probability set \mathcal{M}_i would be straightforward. Namely, given a model M , an observation x_i and a loss function L , the generalized loss function $\mathbb{L}(Y_i, M(x_i))$ would read as

$$\mathbb{L}(Y_i, M(x_i)) = \inf_{p \in \mathcal{M}_i} \mathbb{E}(L(Y_i), M(x_i)) \quad (1)$$

with \mathbb{E} the expectation of the function L of the random variable Y_i having distribution $p \in \mathcal{M}_i$. As Walley's lower previsions [6] provide a subjective interpretation of \mathcal{M}_i without necessitating the assumption that there exists a precise yet unknown probability inside \mathcal{M}_i , I think this kind of modeling is in line with the data uncertainty considered in the paper. I see at least two advantages in doing that:

- It would include both fuzzy and probabilistic uncertainty as particular instances;
- It would allow for more general statements of uncertainty when labeling the data, that could not be properly modeled by fuzzy sets nor probabilities. For instance, over a finite set of classes $\lambda_1, \dots, \lambda_K$, experts could provide lower and upper probabilistic bounds over each class [1] (statements of the kind $\ell_i \leq p(\lambda_i) \leq u_i$) or even provide comparative statements inducing a partial order over the likelihood of observing the classes [4] (statements of the kind $p(\lambda_i) \leq p(\lambda_j)$).

The obvious disadvantages of such an extension to the proposed framework would be an increased computational cost (which could be limited if chosen uncertainty models remain simple enough), and perhaps more importantly a less obvious connection with loss functions, as the uncertainty models could not be reduced to a simple function over the space \mathcal{Y} .

3. The “minimin” approach vs other ones

The approach adopted in the paper can be considered as a decision problem in which an optimistic or a “minimin” approach is adopted. That is, the chosen model (and the associated data disambiguation) is the one that optimizes our decision in the best possible situation. I agree that in the case considered by Eyke Hüllermeier, that is learning a unique best possible model, the “minimin” choice made in the paper is probably the most sensible one, as other decision rules leading to a unique choice (such as minimax or minimax regret) would lead to another unique model likely to not make the best out of the situation. This is quite different from risk analysis or robust optimization problems, where the decision maker wants to prevent him/herself against uncertainty.

However, the alternative mentioned by Eyke Hüllermeier to select a set of optimal models, for instance using the partial ordering induced by the whole distributions $r_M(\alpha)$ in the fuzzy case, or using other approaches such as stating that M_1 is preferred to M_2 when the condition

$$\min \{L(y, M_1(\mathbf{x})) - L(y, M_2(\mathbf{x})) | (\mathbf{x}, y) \in X \times Y\} > 0$$

is satisfied¹, seems to me very interesting. Such approaches would be less conservative than simply applying the extension principle, yet would provide cautious decisions allowing for abstention, as well as a kind of partial disambiguation, as not all replacements would have the same status with respect to the set of possibly optimal models. As the unique optimal model retained in the paper would still be included in this set, one could then still decide to use it in order to obtain a unique model and disambiguation if needed. Of course, the challenge of adopting such a view would be to come up with computationally efficient methods, especially in large problems.

4. The disambiguation process

In the interval or set-valued data case, the data disambiguation process provides a (unique) precise disambiguation or selection, which is a nice result. However, in the fuzzy data case, the disambiguation process currently ends up with a gradual element (number in the case of real-valued data), and it is not clear in the paper how one can go from this gradual element to a crisp and precise replacement (this remark extends to Equation (1), where the disambiguation would be the selected probability distribution). I may be wrong, but in the paper there seems to be no proposed means to obtain a unique precise disambiguation from fuzzy data. Should we select a given alpha-cut? The average value of the gradual element (if such an average can be defined)? The prediction provided by the optimal model and included in the fuzzy set support?

One alternative to obtain such a disambiguation using ideas similar to those of this paper would be to search such a precise disambiguation together with an optimal

¹This is close to the maximal decision criterion [5] used in imprecise probability theory.

model, penalizing models leading to unlikely disambiguations. For instance, if $X \times Y$ are the fuzzy data, we may define an optimal disambiguation x^*, y^* and model M^* as

$$(x^*, y^*, M^*) = \arg \min_{\substack{x, y \in X \times Y \\ M \in \mathbf{M}}} \mathcal{R}_{emp}(M) + \lambda(1 - \mu_{X \times Y}(x, y))$$

where $\mu_{X \times Y}(x, y)$ is the membership value of the disambiguation (x, y) (obtained, e.g., through the extension principle) and where $\lambda \geq 0$ is a parameter that determines how much choosing unlikely disambiguations is penalized. Provided $\mu_{X \times Y}$ is obtained by the extension principle, this would probably correspond to pick a suitable degree α (induced by the value of λ) in the paper method, and then use the set-valued procedure on the α -cut to obtain the optimal model. The above idea could also be extended to more general uncertainty models, replacing μ with a corresponding uncertainty measure (i.e., plausibility measure, upper probability, ...).

5. Likelihood and loss

I think the comparison made with Denoeux's approach sheds a very interesting light on this latter approach, and shows one possible interpretation of it. However, beyond the interpretation in terms of loss functions, I think (but Eyke Hüllermeier or Thierry Denoeux may disagree on that) there is an essential philosophical difference between the two: that while Eyke Hüllermeier's approach includes a subjective decision component in its method through the loss function, Thierry Denoeux's approach tries to obtain what I would call an "objective" description of the data population.

Indeed, Thierry Denoeux's approach aims at estimating the parameters θ of a parametric distribution P_θ describing the data, the distribution P_θ not depending on a particular decision maker. Once this is done, any loss function can be plugged to the model P_θ . On the contrary, Eyke Hüllermeier's approach will lead to an optimal predictive model that may change when the loss function is modified, hence may vary between decision makers.

I think both views offer advantages and disadvantages. If one wants to separate the process of inference from the process of decision (because the loss function to be used is not yet known, or for other reasons [6, Sec. 1.5.]), then I would favor Denoeux's approach. If the loss function (or a good approximation of it) is available and one wants to provide (unique) optimal decisions, then I agree with Eyke Hüllermeier that his approach is likely to make the best out of the available information, and is more adapted in this case.

References

- [1] L. de Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
- [2] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.

- [3] E. Hüllermeier. Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. *International Journal of Approximate Reasoning*, This issue, 2013.
- [4] E. Miranda and S. Destercke. Extreme points of the credal sets generated by elementary comparative probabilities. In *ECSQARU*, pages 424–435, 2013.
- [5] M. Troffaes. Decision making under uncertainty using imprecise probabilities. *Int. J. of Approximate Reasoning*, 45:17–29, 2007.
- [6] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.