International Journal of Approximate Reasoning 55 (2014) 1583–1587

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Discussion

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

Luciano Sánchez

Universidad de Oviedo, Departamento de Informática, 33204 Gijón, Asturias, Spain

A R T I C L E I N F O

Article history:
Received 30 January 2014
Received in revised form 3 April 2014
Accepted 4 April 2014
Available online 23 April 2014

Keywords:
Imprecise data
Fuzzy data
Machine learning
Loss functions
Classification
Regression

A B S T R A C T

The paper by Eyke Hüllermeier introduces a new set of techniques for learning models from imprecise data. The removal of the uncertainty in the training instances through the input–output relationship described by the model is also considered. This discussion addresses three points of the paper: extension principle-based models, precedence operators between fuzzy losses and possible connections between data disambiguation and data imputation.

1. Introduction

The paper [2] describes an appealing method for learning models from imprecise data that improves extension principle-based approaches. The empirical loss of a model on interval-valued datasets is defined as the lowest loss over all the possible crisp instantiations (selections) of the uncertain items in the training data. The loss function of a model with fuzzy data is defined as an average over the different level cuts of the data. The model with a best empirical loss is searched for, thus a minimin criterion is adopted. It is shown that this strategy is related to the optimization of certain loss functions used in machine learning.

This discussion focuses in on three particular aspects of the paper where further developments may be possible: extension principle-based models, the use of the aforementioned minimin criterion and possible links between data disambiguation and data imputation.

2. Application of the extension principle in the context of learning from data

In this section, a nuanced discussion of Prof. Hüllermeier’s comments about the use of the extension principle in the context of machine learning is given. We agree that a direct generalization of standard machine learning techniques to imprecise data via the extension principle will be troublesome in practice. However, if a few minor modifications are introduced, this kind of generalizations can still be useful.

DOI of original article: http://dx.doi.org/10.1016/j.ijar.2013.09.003.
E-mail address: luciano@uniovi.es.

http://dx.doi.org/10.1016/j.ijar.2014.04.008
0888-613X/

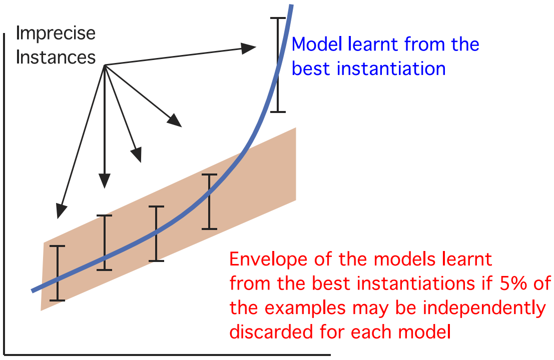

Fig. 1. If there are outliers in an imprecise dataset, all instantiations include them. If a small percentage of the training examples may be removed for each instantiation, the combination of the outputs of the models learned from these incomplete instantiations might be preferred to the output of any single model learned from the complete set.

Many uncertain datasets contain outliers, understood as imprecise items that do not fulfill the assumptions of the model for any of their possible instantiations (selections). In this case, all crisp selections of the dataset may in turn contain outliers. Apart from that, if some items in the imprecise dataset had an abnormally high uncertainty, artificial outliers might also be made up for certain instantiations. This might pose a problem with universal approximators, that might try to learn these extreme values.

A learning algorithm which optimizes a regularized loss function could be used to prevent this. Alternatively, one might also conceive a definition of an empirical loss where some of the training instances that are furthest from the model are ignored (see Fig. 1). A different model is fitted to each selection, thus these items that are discarded will not belong to the same imprecise instances. As a consequence of this, such an empirical loss function cannot always be computed by removing the outliers of the dataset and then applying the extension principle, neither it is equivalent to the application of the extension principle to a regularized loss function.

This idea was introduced in Refs. [6] and [7], where models for precise and imprecise data were proposed that do not produce a point estimate but a crisp or fuzzy set that approximates the union of all the outputs of the models learned from these partial instantiations. In this respect, while we agree that the sentence "We argue, however, that the application of the extension principle is not very meaningful in the context of learning from data" is mostly true, we also think there may be cases where a worthwhile application of the extension principle is possible. Let us remark that the concept of instantiation was not referred to as such in these works, however in the following of this section an effort has been made to cast the ideas introduced in [6] and [7] into the theoretical framework introduced in the paper being discussed.

As mentioned, algorithms [6] and [7] produce crisp or fuzzy families of models, respectively determined by crisp or fuzzy subsets of the parametric space. We will denote them by $\overline{P} \subset \Theta$ or $\widetilde{P} \in \mathcal{F}(\Theta)$, respectively. The fuzzy case is the most general of these two and it is summarized here. Given a crisp input $\boldsymbol{x}$ and a fuzzy subset of the parameter space $\widetilde{P}$, the combined output of the model family $\mathrm{M}_{\widetilde{P}}(\boldsymbol{x})$ is a fuzzy set with membership

$$\mu_{\mathrm{M}_{\widetilde{P}}(\boldsymbol{x})}(\boldsymbol{y}) = \sup\{\mu_{\widetilde{P}}(\theta) \mid \mathrm{M}_\theta(\boldsymbol{x}) = \boldsymbol{y}\} \tag{1}$$

and given a fuzzy input $\widetilde{X}$,

$$\mu_{\mathrm{M}_{\widetilde{P}}(\widetilde{X})}(\boldsymbol{y}) = \sup\{\mu_{\widetilde{P}}(\theta) \wedge \mu_{\widetilde{X}}(\boldsymbol{x}) \mid \mathrm{M}_\theta(\boldsymbol{x}) = \boldsymbol{y}\}. \tag{2}$$

Let ALG be a learning algorithm that builds a model $M_{\widehat{\theta}}$ given a crisp dataset $\mathcal{D} = \{\boldsymbol{z}_i\}_{i=1}^N$, with $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)$. The model $M_{\widehat{\theta}}$ is defined in turn by an estimation $\widehat{\theta}$ of the value of the parameter:

$$M_{\widehat{\theta}} = \mathrm{ALG}(\mathcal{D}). \tag{3}$$

Given a fuzzy dataset $\mathbb{D} = \{\widetilde{Z}_i\}_{i=1}^N$, with $\widetilde{Z}_i = (\widetilde{X}_i, \widetilde{Y}_i)$, a fuzzy estimate $\widehat{P}_0$ of $\widetilde{P}$ could be thought of that only involves complete instantiations,

$$\mu_{\widehat{P}_0}(\theta) = \sup\{\mu_{\mathbb{D}}(\mathcal{D}) \mid \mathrm{ALG}(\mathcal{D}) = M_\theta\} \tag{4}$$

where $\mu_{\mathbb{D}}(\mathcal{D}) = \min_{i=1}^N \mu_{\widetilde{Z}_i}(\boldsymbol{z}_i)$. This kind of estimation we agree with Prof. Hüllermeier that it is not a good approach because more often than not we would end up with a set of predictions which is highly nonspecific.

On the contrary, an estimation restricted to certain subsets of the instantiations may be useful. In short, $\widetilde{P}$ will be determined as the set that minimizes a functional of the parametric model M_θ, which is not derived from an extension principle-based generalization of the crisp loss. A set $\widetilde{P}$ is searched for such that the average nonspecificity of the (extension