International Journal of Approximate Reasoning 55 (2014) 1609–1613

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Discussion

Rejoinder on "Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization"

Eyke Hüllermeier

Department of Computer Science, University of Paderborn, Germany

ARTICLE INFO

Article history:
Received 3 April 2014
Accepted 7 April 2014
Available online 1 May 2014

Keywords:
Machine learning
Imprecise data
Extension principle
Disambiguation
Loss function
Classification

ABSTRACT

This is a short note in which I reply to the comments and suggestions made by Luciano Sánchez, Sebastian Destercke and Didier Dubois on my article "Learning from Imprecise and Fuzzy Observations: Data Disambiguation through Generalized Loss Minimization".

1. Introduction

First of all, I would like to sincerely thank Luciano Sánchez, Sebastian Destercke and Didier Dubois for their interest in my work and for studying my paper [3]. I am very grateful for their commentaries on my article, in which they do not only provide a number of thought-provoking remarks but also suggest very interesting directions for future work.

2. Reply to Luciano Sánchez

Following the structure of his own document [4], I will reply to the aspects addressed by Prof. Sánchez in exactly the same order: The application of the extension principle in the context of learning from data, the combination of model induction and data disambiguation, and the idea of exploiting approaches to data disambiguation for the purpose of data imputation in cases of missing observations.

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

The main points that I extract from the comments of Prof. Sánchez in this section are the claims that (i) my approach of learning from imprecise or fuzzy data might be prone to the problem of over-fitting the training data, especially in cases

DOI of original article: http://dx.doi.org/10.1016/j.ijar.2013.09.003.
DOIs of comments: http://dx.doi.org/10.1016/j.ijar.2014.04.008, http://dx.doi.org/10.1016/j.ijar.2014.04.014, http://dx.doi.org/10.1016/j.ijar.2014.04.002.
E-mail address: eyke@upb.de.

http://dx.doi.org/10.1016/j.ijar.2014.04.012
0888-613X/

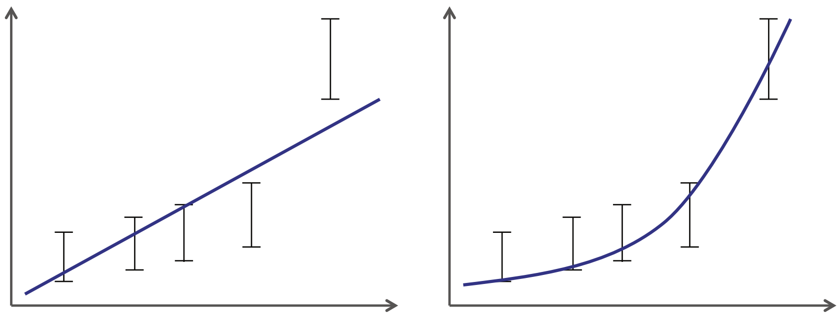

Fig. 1. Reproduction of the example of Prof. Sánchez: Fitting the data with a simple linear model (left) versus fitting a mode complex, non-linear function (right).

where the data contains outliers, and (ii) that methods based on the use of the extension principle might be more robust in this regard. I do not completely agree with these points and would like to put them in perspective.

To begin with, I would like to make a general remark on the first point. My approach is not a concrete machine learning method but should be seen as a kind of meta learning method. Just like the extension principle, it allows for generalizing an existing statistical or machine learning method from the case of precise to the case of fuzzy data. In general, the sensitivity toward outliers is more a property of the concrete method and less of the meta method. For example, my approach can be applied to linear regression with L_1 loss minimization in the same way as to non-linear regression with L_2 loss minimization. Needless to say, the latter will be much more sensitive toward outliers than the former.

In this regard, Prof. Sánchez is making some assertions, at least implicitly, that might be somewhat misleading. First, I would like to stress that my approach is not restricted to *empirical risk minimization* (ERM), an induction principle that is indeed prone to over-fitting. Instead, as mentioned in the first part of Section 4 of my article, regularization can be incorporated in a straightforward manner. It was only for reasons of simplicity and ease of exposition that the whole approach was developed for the ERM.

Second, the discussion as well as the illustration in Fig. 1 by Prof. Sánchez seem to suggest that the models produced by my approach are somehow passing through all the output intervals or fuzzy sets, which is certainly wrong (in particular, note that the intervals may reduce to single points). The approach minimizes a generalized loss on the training data but will usually not be able to achieve a total loss of 0. The example in Fig. 1 by Prof. Sánchez does not properly illustrate the effect of over-fitting either. In fact, given the data shown in this example, there are two cases that need to be distinguished.

- First, the true dependency might be simple, for example a linear regression function. In this case, taking the class of (affine) linear functions as an underlying model class (denoted **M** in my article), the model learnt may look like the one shown in the left panel of Fig. 1 (in which the example of Prof. Sánchez is reproduced). Here, the rightmost observation could indeed be seen as an outlier. It's influence will then depend on the underlying regression method, especially on the loss that is minimized (e.g., L_1 versus L_2), but in any case, this influence will be limited.
- Second, the true dependency might be a more complex, non-linear one. In that case, the solution may indeed look like the one in the right panel of Fig. 1 and resemble the one shown by Prof. Sánchez. In that case, however, why should the rightmost observation be seen as an outlier? In fact, assuming a complex dependency, the set of observations is completely plausible, and the model fit to the data makes perfect sense.

Besides, of course, nothing speaks against the suggestion of Prof. Sánchez to completely remove potential outliers from the data before fitting a model. Beyond any doubt, my approach can be generalized in this direction, and can be combined with such kind of pre-processing techniques. In fact, it is even very suitable for that purpose: Fuzzifying data is one way of modulating their influence on model induction. For example, if the rightmost observation in Fig. 1 appears to be an outlier, then it's influence can be reduced by widening the corresponding interval (or fuzzy set). Thus, the approach offers the possibility to smoothly vary between fully using an observation and completely ignoring it.

Regarding the alternative approach outlined by Prof. Sánchez, I certainly agree that looking at more than a single data instantiation could be useful (for example, with regard to producing confidence intervals or, more generally, reliable estimates). Still, the application of the extension principle will come with the problem that model selection and data instantiation are not fully attuned to each other, which is in conflict with my idea of *data disambiguation*.

Moreover, as far as I understand the description of Prof. Sánchez, his model's robustness toward outliers should be attributed less to the use of the extension principle but more to a built-in mechanism for handling outliers. In fact, a model is searched under the constraint that *the fraction of covered instances exceeds a given threshold* (while the remaining ones, which might be potential outliers, do not need to be covered). However, there are several issues with an approach of that kind, notably the following: