International Journal of Approximate Reasoning 52 (2011) 1229–1231

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Editorial

Special issue: Handling incomplete and fuzzy information in data analysis and decision processes

This issue contains a selection of papers presented at the joint IFSA-EUSFLAT Conference in Lisbon, July, 20–24, 2009. This meeting put together two bi-annual events dealing with fuzzy sets and their applications, namely the World Congress of IFSA (The International Fuzzy Systems Association) and the European Conference of EUSFLAT (The European Society for Fuzzy Logic and Technology).

The topic of the special issue is centered on the handling of incomplete information in data analysis and decision processes, using fuzzy sets, possibility theory, and fuzzy measures. It is well-known that uncertainty can be caused by the variability of observations of phenomena, or may be due to incomplete or contradictory information. In the latter case, subjective probabilities have been developed to cope with the lack of statistical evidence supporting a distribution. However the use of unique probability distributions is not so satisfactory an approach when facing incomplete knowledge. Set-valued approaches like interval analysis and their generalisation using possibility theory look more convincing as less biased. Then uncertainty is represented by fuzzy sets. The collection of papers here enclosed witnesses the usefulness of set-valued methods and possibility theory-based representations of uncertainty.

The two first papers illustrate connections between statistics and fuzzy sets. Gilles Mauris unifies, by means of possibility theory, many old techniques independently developed in statistics for one-point estimation, relying on the idea of dispersion of an empirical distribution. The efficiency of different estimators can be compared by means of fuzzy set inclusion applied to possibility transforms of probability distributions. This unified approach does not presuppose a finite variance. Pedro Teran focuses on the extraction of values considered to be centers of distributions, understood as values which make an event probable if it contains them. This notion is intuitively not Boolean, and the author proposes a general approach to measure degrees of centrality, relying on fuzzy set theory to build an inclusion index representing this notion. Again this approach unifies many notions of central points known in statistics.

The two subsequent papers have to do with regression and classification of imperfect data. The contribution of Boukezzoula and colleagues deals with the extension of linear regression to interval data. This work is in the line of so-called possibilistic regression initiated in the 1980s by Hideo Tanaka and his students. The purpose is to fit a linear model whose coefficients are intervals to a partially observed set-valued function which assigns an output interval to each precise numerical input. The idea is that the interval-valued model should enclose the interval data as tightly as possible. One original feature of the proposed approach is the use of the *midpoint-radius* representation of the intervals, which allows to reason about the trend of the intervals and their widths in a separate way. Interval linear models are monotonic not only in their trends but in terms of output width as well, a feature that may cause difficulty in the regression process; the authors then introduce a piecewise linear interval regression model.

The paper written by Ana Colubi and colleagues considers the issue of classifying fuzzy data. They assume each class is a class of fuzzy sets. Authors work with a space of random fuzzy sets, understood as metric-space-valued random elements. This space is equipped with a scalar distance between fuzzy sets. The problem is to estimate the probability distribution attached to each class. Once this is done, Bayes rule is used to compute the probability of membership of a fuzzy object to each class. In order to come back to classical probability calculations, the fuzzy object is replaced by a ball of fixed radius around the fuzzy object. In order to avoid computational difficulties linked to the density estimation problem in the context of random fuzzy sets, the authors propose an alternative approach avoiding this density estimation step. The success of the method relies on the proper choice of the radius of the ball around a fuzzy object to be classified. The techniques are empirically validated on two examples, one pertaining to perceptions of the length of a ruler modelled by trapezoidal fuzzy sets on a continuous scale, one on a flood analysis problem where flood heights are reported by visual inspection.

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2011.08.006

What is striking in the fuzzy regression and classification literature is that the epistemological status of intervals or fuzzy intervals is not always clear, despite the often impeccable technical machineries at work in the proposed methods (for instance, as in the two papers on this topic in the special issue). An interval, let alone a fuzzy interval, may either represent a real object (a road segment in some place) or stand for incomplete knowledge about an ill-known point-value. These two points of view correspond to what we have elsewhere called the *ontic* view and the *epistemic* view of sets, respectively [2]. The optimisation criterion used by Boukezzoula et al., minimizing discrepancies between the interval data and the enclosing interval-model, makes sense for both interpretations of intervals. However, the use of an interval-valued linear model is more questionable in the epistemic view, beyond its limited expressiveness. If the interval-valued model represents an imprecisely known precise model, it is not clear why the imprecision of the output should linearly depend (actually, should depend at all) on the objective values of the input. Moreover, the problem addressed in the paper of Boukezzoula et al. has little to do with the study of the impact of the ill-knowledge of point-valued data (modelled by intervals) on the imprecision of the standard linear regression model one would have computed, had the data been precise. The same questioning can be raised for the paper by Colubi et al. where the random fuzzy sets are handled by means of functional analysis and treated as objective entities, since a scalar distance between them is used. However, in the application examples it is not clear that the fuzzy intervals obtained from observations should always be considered as ontic. It all depends on whether *perceptions* are considered as data per se (as done in the paper), or if one is interested by the actual value imperfectly accounted for by the perception representation (the actual length of the ruler, or the actual height of the flood at a given place). Dealing with the latter requires a form of sensitivity analysis. In this case, fuzzy random variables are more naturally interpreted as epistemic fuzzy sets of random variables than as ontic random fuzzy sets [1].

The paper by Fan and colleagues deals with incomplete data tables on ordered scales and extends the rough classification method developed by Greco, Matarazzo and Słowinski to imprecise data. Interestingly, they outline two methodologies to handle this problem. In the first approach, they interpret incomplete data probabilistically using Laplace indifference principle, and compute the degree of preference of one object over another one using a statistical preference index. In the second approach, they take ignorance for granted and model incomplete data by means of possibility distributions, hence exploit possibilistic preference indices between objects. In this paper, interval-valued data are clearly interpreted as epistemic sets.

This is also the case with the contribution by Kasperski and Zielinski. These authors present a full-fledged extension of a class of combinatorial optimisation problems (bottleneck problems) to incomplete and fuzzy data. Quite at the opposite to the regression and classification papers appearing in the special issue, the main concern of the authors here is to model and compute the impact of the imprecision of the data on the notion of optimal solution. Handling fuzzy interval data leads them to rigorously define degrees of necessary optimality and degrees of possible optimality that have to be optimised in place of the usual criterion in the precise data situation. This is done in the setting of ordinal possibility theory, where possibility degrees just reflect closeness to plausible values and no statistical interpretation is assumed. The paper also relates these decision criteria under possibilistic uncertainty to the minimal regret criterion, so as to reach robust possibly optimal solutions, when necessarily optimal ones do not exist. This is achieved by defining a fuzzy set standing for a fuzzy goal and computing solutions maximising the necessity of reaching this fuzzy goal, which is the counterpart in possibility theory to both maximin and expected utility criteria [3].

The two last papers deal with aggregation operations. Gagolewski and Grzegorzewski study ordinal aggregation operations generalising the ordered weighted maximum and the *h*-index for the purpose of improving bibliometric indices of the latter kind. The role of possibility distributions is to supplement the proposed approach with a dynamic or predictive component allowing to compare two scientists having the same static index value in terms of the likelihood of one becoming better than the other in the future. Interestingly, this method makes sense for evaluating any machine producing items whose value is measured by their popularity or generated profit (e.g. how often they are bought, instead of cited); it highlights the current "productivist" view of the research activity.

The last paper of the special issue is of a more theoretical nature, and investigates properties of set-functions involved in aggregation operations as well as in game theory and in information fusion. The main focus is the design of tools for simplifying the representation of complex set-functions, thus making a trade-off between expressiveness and complexity. Narukawa and Torra introduce so-called *m*-decomposable set-functions. The output of the set function on a given set is computed with respect to a partition of the universe, by some aggregation function merging the evaluations of the contribution of each partition element to the set under concern. The authors study formal links between this construction and *m*-dimensional distorted probabilities introduced in the literature of decision theory.

The guest-editor wishes to thank the authors of this special issue for submitting interesting papers, and for their patience and good spirit during the reviewing process. Many thanks go to the Editor-in-Chief of this journal for welcoming this special issue and for his constant, active and friendly support during the reviewing process. The guest-editor also expresses his gratitude to Uzay Kaymak, co-chair of the IFSA-EUSFLAT 2009 conference, for suggesting this special issue and for further discussions on its content.

References

[1] I. Couso, D. Dubois, On the variability of the concept of variance for fuzzy random variables, IEEE Transactions on Fuzzy Systems 17 (5) (2009) 1070–1080.

[2] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: making sense of fuzzy sets, Fuzzy Sets and Systems, in press, doi:10.1016/j.fss.2010.11.007.
[3] D. Dubois, H. Prade, R. Sabbadin, Decision theoretic foundations of qualitative possibility theory, European Journal of Operational Research 128 (2001) 459–478.

Didier Dubois
IRIT, CNRS and Université de Toulouse, France
E-mail address: dubois@irit.fr

Available online 3 September 2011