ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Approximate Reasoning
39 (2005) 115–121
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Editorial

Credibility via imprecise probability

1. Introduction

This special issue of the *International Journal of Approximate Reasoning* grew out of the *Third International Symposium on Imprecise Probabilities and Their Applications* (ISIPTA '03), [1] held in Lugano, Switzerland, in July 2003. The origins of the biennial ISIPTA meetings date back to 1999. At that time a group of people joined their efforts in the organization of the first meeting, ISIPTA '99, under a specific impulse: to make it possible for scientists in different disciplines to discuss their views about imprecision in theories of probability. Indeed, imprecision in probability has been studied almost independently in many different fields, such as Computer Science, Economics, Engineering, Psychology, Statistics, just to name a few. Each of these started its own traditions and especially its own language to talk about imprecise probability. Better communication seemed to be a necessary step in order to make the various approaches to imprecision in probability theory benefit from each other, and perhaps eventually develop some common ground. ISIPTA '99 actually did witness lively and healthy debates on these matters, and formed the basis for an interdisciplinary community which is now very actively involved in advancing and promoting a generalized view of probability theories. In February 2002, this community saw the birth of the *Society for Imprecise Probability Theory and Applications*. The Society [2] manages the organization of the ISIPTA meetings, nowadays a well-established international forum for discussing imprecise probabilities, and has recently organized also the first *summer school on imprecise probabilities*, [3] of which many more will hopefully follow.

Imprecise probability is really a heterogeneous field, as it is testified by the papers presented at ISIPTA '03. A total of 44 papers were presented that covered a wide range of topics, including: new model based inference with imprecise probabilities;

[1] See http://www.sipta.org/isipta03
[2] See http://www.sipta.org
[3] See http://www.idsia.ch/~zaffalon/events/school2004/school.htm

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2004.11.001

computations and foundations of inference with imprecise probabilities; applications of imprecise probabilities in engineering, finance, and medicine; connections with graph theory, belief functions, and fuzzy random variables; and the introduction of new principles and tools for decision theory. ISIPTA '03 included also three invited contributions from Terrence L. Fine, Irving J. Good, and Patrick Suppes, and five invited tutorials [4] (Jean-Marc Bernard: "Imprecise Dirichlet model for multinomial data," Gert de Cooman: "A gentle introduction to imprecise probability models and their behavioral interpretation," Fabio G. Cozman: "Graph-theoretical models for multivariate modeling with imprecise probabilities," Charles F. Manski: "Partial identification of probability distributions," Sujoy Mukerji: "Imprecise probabilities and ambiguity aversion in economic modelling").

2. Imprecise probabilities

It should be clear from the Introduction that the term 'imprecise probability' covers a lot or areas. 'Probability' is used for any theory of randomness or uncertainty; 'imprecise probability', as a generic term for the many mathematical or statistical models that measure chance or uncertainty without sharp numerical probabilities. These models include, but are not necessarily limited to, belief functions, Choquet capacities, comparative probability orderings, convex sets of probability measures, fuzzy measures, interval-valued probabilities, possibility measures, plausibility measures, and lower and upper expectations or previsions. It is often said that these imprecise probability approaches are needed in inference problems where the relevant information is scarce, vague or conflicting, and more generally that they have to do with modelling partial ignorance, incompleteness, indeterminacy and indecision. Actually, many of these aspects can be regarded as different expressions of a specific characteristic of imprecise probability that I will emphasize in the following: imprecise probability offers the opportunity to make probability models less fragile, and hence more credible, by acting on the relationship between assumptions and conclusions. Here 'assumptions' is used in a loose way to generically express the restrictions imposed on our models on the way to abstraction. These assumptions determine, for example, the language used to represent the evidence. It seems most important, though, to consider that the connection between assumptions and conclusions is intimately related to the scientific method itself, as scientific activity is mostly concerned with drawing conclusions from assumptions (and evidence). In this light, assumptions are of a foundational nature in Science; they are the pillars that support our conclusions. These pillars need to be robust in order to support credible conclusions; and the key point is that the way to robustness passes through designing assumptions so as to make them tenable.

Manski [5] expresses a similar concept very nicely with his *law of decreasing credibility*: "The credibility of inference decreases with the strength of the assumptions

[4] The technical papers, as well as the invited contributions and tutorials, are publicly available from the web site of the conference.

[5] See Manski's tutorial in this issue.

maintained." Of course, this is not to say the weaker assumptions the better, rather, the stronger the better, but only up to the point where they remain tenable.

This standpoint places a strong emphasis on the role of assumptions, in the sense that it regards a rigorous design of assumptions as a necessary component of a rigorous scientific approach. Rigor requires assumptions to be critically discussed, and their implications to be deeply understood. But, most of all, it requires a formal framework and a well-developed body of mathematical results that enable one to state and work with tenable assumptions. Without both, assumptions might have to be made purely for mathematical convenience, which is clearly not desirable.

Imprecise probability appears to be a candidate framework to state and work with tenable assumptions. The key feature of imprecise probability in this respect is their expressivity, which allows one to represent a full range of degrees of knowledge, typically from complete ignorance to deterministic knowledge, passing through traditional uncertainty representations. How this can be done in practice can be better shown with some examples. The papers in this special issue are well suited to serve this purpose, given also the variety of contexts they represent, and their application-oriented flavor. The next section will give a brief summary about the papers from this very perspective.

3. The papers

This issue collects extended versions of six papers and three tutorials presented at ISIPTA '03. ISIPTA '03 emphasized three main themes: inference from data, algorithms, and applications. The same themes have been used as guidelines in the selection of the articles for this special issue, in order to emphasize the practical use of imprecise probabilities mentioned in the previous section.

Bernard's tutorial on the *imprecise Dirichlet model* (IDM) opens the issue. The IDM is a model of statistical inference from multinomial data for the case when substantial prior knowledge on the process producing the data is not available. The IDM can be regarded as a generalization of the popular Bayesian learning method that makes use of Dirichlet priors. The IDM relaxes a specific assumption of its precise counterpart, namely, that a prior state of ignorance can be represented satisfactorily by a single prior density. The tutorial argues, from a number of viewpoints, that absence of prior knowledge can be better modelled by sets of prior Dirichlet densities. The machinery of Bayesian learning is then extended in a straightforward way to manage the multiple priors, producing in general lower and upper probabilities and expectations. The tutorial illustrates also some existing applications of the IDM, while making it clear that there remains a great potential for applications to be explored. In artificial intelligence, for instance, inference from categorical data is a major task. One example is the inference from data of model probabilities for Bayesian networks. Sometimes there is a problem of credibility with respect to these probabilities, especially for complex networks and limited sizes of data. The IDM might help in addressing this problem, since by exploiting the imprecision represented by lower and upper probabilities, it produces credible probabilities for any size of the learning

data. More generally speaking, the tutorial is intended as a way to disseminate the benefits of using of multiple priors in the artificial intelligence community.

If Bernard's tutorial in on the side of multiple priors, the subsequent tutorial by Manski can be intended as a way to work with multiple likelihoods. Multiple likelihoods arise in the case when the observed data are consistent with more than one likelihood, as with missing data. In this case, different likelihoods follow from different ways to replace the missing values with actual values. The by far more popular way of dealing with missing data is to assume some kind of *missing at random* (MAR) condition, which basically states that the mechanism that produces the missing data is not selective. MAR makes it relatively easy to deal with missing data as the inference is then allowed to be based only on the data that are present. Furthermore, MAR allows everything to be treated within a precise probability setting. These advantages are unfortunately dwarfed by two serious problems: MAR is a strong assumption in general and, also, MAR cannot be tested as there is no statistical test for it. Clearly, there is a serious risk of arbitrariness in assuming MAR, and hence a serious risk of producing misleading conclusions. In the last fifteen years, Manski's research has built up an entire framework for dealing with missing data without assuming MAR, and that naturally produces imprecise probabilities. Remarkably, that framework makes it possible in many cases to produce meaningful conclusions also when there is real ignorance about the missingness mechanism. Manski also discusses ways to enforce assumptions about the missingness mechanism in order to produce stronger, while still credible, conclusions. Manski's tutorial should serve as an introduction to a credible way of dealing with missing data, which can be deepened by means of the cited references, and in particular by means of Manski's recent book, which is the source of inspiration for the tutorial.

The third article is the last tutorial of the special issue: Cozman's introduction to graphical models for imprecise probabilities, focused mostly on *credal networks*. Credal networks can be regarded as an extension of Bayesian networks to imprecise probabilities. In the context of probabilistic expert systems, credal networks relax the assumption that expert knowledge must be modelled by a precise probability distribution. More credibly, expert knowledge is assumed to be vague to some extent, and to be more faithfully captured by imprecise statements, such as constraints on probability values. Credal networks provide a means to represent this kind of knowledge and to produce conclusions that are fully consistent with it, similarly to what can be done in the precise case with Bayesian networks. Credal nets are also an important tool in a statistical setting. Indeed, statistical approaches based on multiple priors, or on multiple likelihoods, naturally produce imprecise probabilities. If we want to join the power of graphical models with those statistical methods, we need some kind of imprecise graphical model, such as credal nets. Cozman's tutorial reviews the most important aspects of credal networks, such as the interpretation of the model, some special cases of the general definition, and algorithms to compute with credal networks.

Kriegler and Held then present a paper on the application of imprecise probability to the environmental problem of assessing climate change. Of course this is a very important topic, and it is also a way to illustrate clearly how imprecision comes natu-

rally into place in an application. Indeed, the authors show that at present there is very little information about many important quantities that have the potential to severely influence climate change. On the other hand, it is important to make reliable predictions as scientific results will (hopefully) be the basis of future policies to minimize the risk of a serious drift in environmental parameters. In order to avoid overconfident conclusions, the authors collect a set of methods that allows ignorance about key parameters to be included in a simple climate change model. Within the limits of the chosen modelling approach, they show that the uncertainty about the temperature increase in the 21st century is large. Despite this, they find that it is very unlikely that the warming will be less than two degrees Celsius. This is an important point as staying below the threshold of two degrees Celsius is a frequently discussed climate protection target in order to avoid dangerous climate change. The authors' approach turns then out to be useful to indicate which variables account for which portion of the uncertainty.

In a sense, the environmental application described by Kriegler and Held favors approaches based on imprecision, for its highly uncertain nature. The next paper by Capotorti is a nice example to show that imprecision comes naturally into place also in more circumscribed applications, especially with cautious modelling approaches. The paper focuses on two medical applications, for which domain knowledge is relatively scarce. Capotorti refrains from making strong assumptions in order to produce determinate conclusions. Rather, he illustrates an approach that starts by making weak assumptions and strengthens them little by little. Initially, domain knowledge is formalized by simple constraints on probabilities. In both problems, the formalization naturally produces partially indeterminate conclusions. In some cases these are not informative enough, as the related intervals are too wide. Of course, this may happen when we accept that our models can be imprecise, and it is a useful indicator that we are missing relevant knowledge in order to draw any serious conclusion. In order to maintain credibility, the author then carefully considers some further assumptions, which eventually produce strong enough conclusions. Capotorti's approach in interesting at least for two reasons. First, it shows that imprecise probability models are sometimes better developed in a step by step fashion: the results of the first modelling attempts are used to understand whether stronger assumptions, and which ones, should be considered next. Secondly, the paper proposes some so-called *structural* assumptions to strengthen conclusions in an imprecise probability setting. These are relatively general, so that they might profitably be considered in a variety of applications.

The next three papers of the issue have a more computationally oriented flavor. In the first, Abellán and Moral investigate the application to pattern classification of a measure of entropy for sets of distributions (also called *credal sets*). They do this in the setting of multiple priors, by extending a well known classification tree to credal sets originated by the imprecise Dirichlet model. The results are important and twofold. First, the authors show empirically that trees constructed with the new measure of entropy automatically avoid overfitting problems; this, in turn, enables the new trees to perform consistently better than traditional ones with respect to prediction accuracy. Second, the proposed classifier offers the opportunity to do classification

credibly for any size of the learning data. This is achieved by producing set-based classifications (i.e., partially indeterminate classifications) for the objects that are more difficult to classify, given the knowledge available in the learning set. Classifiers with this characteristic are called *credal classifiers*.

In the subsequent paper, De Cooman and Troffaes wonder whether the popular method of *dynamic programming* may be extended to settings where uncertainty on the gain function is, perhaps more realistically, described by imprecise probabilities. The authors formalize elegantly the problem in a very general framework, which eventually leads to interesting conclusions. For example, they show that approaches based on the *maximin* decision criterion cannot generally be extended to the new setting, as they need some specific assumptions. Considered that the maximin has received considerable attention so far, this appears to be an important result, as it allows other researchers to be aware of the inherent assumptions made by using such a criterion. It is also worth remarking that other decision criteria, such as *maximality* and *robust optimality*, are shown to be always extendible, which contributes to these being regarded as criteria worthy of investigation.

Exact and approximate algorithms for credal networks are the subject of the next paper by Da Rocha and Cozman. The authors have been very active in this field, with much work aimed at making credal networks a more and more practical modelling tool. Indeed, it is well known that computing with credal networks is generally harder than with Bayesian networks, so that substantial efforts must be devoted in order to be able to compute with the former, especially in complex domains. In particular, the authors present an exact algorithm that works well with relatively small networks, and an approximate one for larger models. It is worth noting that the latter produces *outer approximations*, which are usually of most interest when working with imprecise probabilities. Outer approximations produce intervals of probabilities, or expectations, that enclose the exact ones. This is an opposite approach with respect to *inner approximations*, which are enclosed by the exact intervals, and which generally produce overconfident conclusions. Outer approximations, in other words, always work on the 'safe side' by adding some extra caution to the inference when exact results cannot be obtained. This kind of *algorithmic imprecision* is another important feature of imprecise probabilities, that allows credibility to be maintained despite computational limitations. In the specific case, Da Rocha and Cozman's approximate algorithm is shown to produce significantly better approximations than a previously known one on which it is based, and it is so a promising approach for the concrete use of credal networks.

Pelessoni and Vicig's paper concludes the issue. This paper is somewhat different from the others as it is more focused on foundational rather than applied issues. Yet, it is still motivated by developing credible models, in the following sense. At present, among the more general imprecise probability models there are Peter Walley's *coherent lower previsions*, which are equivalent to closed convex sets of probability distributions. Coherent lower previsions are based on rationality criteria called *avoiding sure loss* and *coherence*. Pelessoni and Vicig argue that tools based on weaker assumptions should be considered in some cases. They propose to this extent a model called *center convex previsions*, which avoid sure loss but are not necessarily coher-

ent. The result is a new very general theory, both for the unconditional and the conditional case, which appears to be very powerful and with substantial relations, and implications, for financial risk measurement.

4. Conclusions

This issue was conceived as a way to disseminate ideas from imprecise probabilities in the community of approximate reasoners. It tries to do so by showing that it is possible to develop, and practically apply, imprecise probability models based on tenable assumptions right now. Of course imprecise probability is also a complex field, with many open problems and controversies. There is the need of a great joint effort to tackle these challenges. I hope that this issue contributes to have more people exploit the wealth of opportunities for research offered by imprecise probabilities, both at the foundational and application level.

Acknowledgments

There are many people I would like to thank, who in some way helped to make this issue possible. First, I wish to thank Jean-Marc Bernard and Teddy Seidenfeld. I had the great opportunity to collaborate with them for the organization of ISIPTA '03, a task that turned out to be both instructive and fun. The ISIPTA '03 organization was always under the watchful eye of the steering committee, which I really wish to thank, in the persons of Gert de Cooman and Serafín Moral. Serafín also managed the electronic procedures for the conference, a time-consuming task for which he volunteered in a very generous fashion. Many thanks also to the eighteen anonymous referees for this issue. The editorial work has greatly benefited from their work, as they all discharged their responsibilities very effectively. The editorial itself has also benefited from one review; thanks again to Gert de Cooman for this, whose comments helped me express my ideas more clearly. Finally, I wish to thank Piero Bonissone very much for always being helpful and flexible in our collaboration for this special issue. This work received partial support by the Swiss NSF grants 20C021-101209 and 2100-067961.

Marco Zaffalon
IDSIA, Istituto Dalle Molle di Studi
sull'Intelligenza Artificiale Galleria 2
CH-6928 Manno (Lugano)
Switzerland
Tel.: +41 58 666 666 5; fax: +41 58 666 666 1
E-mail address: zaffalon@idsia.ch

Available online 8 December 2004