Artificial Intelligence 175 (2011) 1037–1052

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Preferences in AI: An overview

Carmel Domshlak [a], Eyke Hüllermeier [b], Souhila Kaci [c], Henri Prade [d]

[a] Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Israel
[b] Department of Mathematics and Computer Science, Philipps-Universität Marburg, Germany
[c] CRIL, IUT de Lens, Rue de l'Université, SP 16, 62300 Lens, France
[d] IRIT, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France

ARTICLE INFO

Keywords:
Preference
Graphical representation
Logical representation
Aggregation functions
Preference queries
Preference learning

ABSTRACT

This editorial of the special issue "Representing, Processing, and Learning Preferences: Theoretical and Practical Challenges" surveys past and ongoing research on preferences in AI, including references and pointers to the literature. It covers approaches to representation, reasoning and learning of preferences. Methods in AI are contrasted with those in related areas, such as operations research and databases. Finally, we also give a brief introduction to the contents of the special issue.

1. Introduction

Even if the purpose of reasoning is often to support decision making, only since the 1990s has decision theory had much impact on AI, initially in connection with planning under uncertainty (e.g., [1]). The modeling of preferences is a prerequisite for any kind of further decision analysis. It becomes a non-trivial issue as soon as the preferences cannot be expressed in a binary way, distinguishing good alternatives from bad ones, and easily enumerated in terms of an explicit list.

The treatment of human decision problems requires a clear distinction between knowledge (pertaining to the current state of the world) and an agent's preferences among possible states. Mixing binary preferences, easily expressed in logic, with a logical knowledge base leads to 'taking desires for reality'. Knowledge may be pervaded with uncertainty, an issue that has been considered in AI since the emergence of expert systems. In principle, uncertainty may also apply to preferences, but this is less crucial since decision under uncertain preferences is rarely considered. Instead, starting with a set of 'rationality' postulates, the classical framework of Savage's decision theory [2] justifies the probabilistic modeling of the knowledge about the present state of the world, together with a numerical representation of preferences in the form of a value function that precisely assesses the possible results that might be achieved through different actions.

The increasing importance of decision-making to AI has led to a growing focus on the management of preferences [3], especially fostered by the advent of graphical representations [4,5] in the late 1990s, partly inspired by the use of similar representations for knowledge in Bayesian networks. This has let to a series of important workshops [6–11] and to special issues of leading journals [12–14], where other types of representations were discussed as well.

Before presenting the contents of this special issue in Section 5, we start with a brief historical outline in Section 2, where research on preferences in AI is positioned with respect to contributions from operations research (OR) and databases (DB). These fields are especially relevant for AI, although other fields could of course be mentioned, too. In fact, it should be emphasized that preferences is an interdisciplinary topic that can be studied from different perspectives. As an important example, we mention the study of human preferences in psychology, notably in connection with decision making [15,16];

E-mail addresses: dcarmel@ie.technion.ac.il (C. Domshlak), eyke@mathematik.uni-marburg.de (E. Hüllermeier), kaci@cril.fr (S. Kaci), prade@irit.fr (H. Prade).

0004-3702/$ – see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.artint.2011.03.004

see [17] for a survey. Such studies may, and to some extent already did, serve as a source of inspiration for AI research and validation of AI models.

The main research topics in AI are then surveyed in more depth in Sections 3 and 4. While the former is focused on representing and reasoning with preferences, the latter is devoted to the learning of preferences.

2. Preferences in AI and related fields

The representation of preferences has been studied in economics, especially in decision theory and in social choice theory, with further developments and applications in OR, long before AI or database researchers became interested in the topic. Here, we briefly outline from what perspectives the modeling of preferences has been studied, and try to highlight the main characteristics of the approaches developed in these fields. We begin our discussion with economics and OR before considering AI and DB contributions.

2.1. Preferences in economics and operations research

Preferences are central not only to individual decision making, but also to collective decision making, known as social choice, and the study of strategic interactions between agents, the topic of game theory. The formal developments of decision theory, social choice and game theory all emerged in economics around the same time (between late 1940s and early 1950s): [18] for decision theory and game theory, [19] for social choice. These areas now play a huge role in AI: decision making under risk and uncertainty in planning (and especially Markov Decision Processes), and social choice and game theory in most formal studies of multi-agent systems (voting, resource allocation, auctions, etc.) [20]. We now briefly outline what economics and OR have provided in terms of preference modeling, which will serve as reference material for the AI research discussed later.

In decision making under uncertainty, a preference relation between acts is built from a probability distribution over the possible pairs of input and output states and from a utility function assessing the value of each result [21]. In Savage's decision theory [2], one act is preferred to another if its expected utility is higher.

Generally speaking, expected utility can be seen as the prototype decision criterion proposed in decision theory. It may be considered as an instance of the relational modeling of preferences viewed as a conjoint measurement problem [22–24], where a binary (preference) relation is defined between objects described by vectors. Each vector encodes an act by the values of its result when performed in different states of the world in the case of decision making under uncertainty, or lists the evaluations of an alternative according to different criteria in case of multiple criteria decision making, or according to different agents in group decision making. Conjoint measurement theory then looks for conditions under which there exists a numerical representation (possibly unique) of the preference relation in the form of a decision criterion. This type of representation requires preferences to be complete and transitive. Intransitive models have been studied as well [25–27].

A decision criterion in decision making under uncertainty aggregates the values of consequences of an act obtained in different states of the world. What is aggregated in multiple criteria decision making, instead, are numerical satisfaction degrees pertaining to the different criteria that are considered. Different types of scales [28,29] can be used for assessing these satisfaction degrees: ordinal scales where only the ordering of the grades is defined, interval scales where numerical grades are defined up to a positive affine transformation, and ratio scales where the grades are defined up to a multiplicative factor. Depending on the type of the scale, different families of aggregation functions may be used (conjunctions, disjunctions, averages, ordered weighted averages (OWAs) [30,31], ordered weighted conjunctions [32], etc.), and many studies have looked for axiomatic characterizations of these families in terms of properties that are easy to interpret in practice [33–38]. However, scoring functions cannot represent all preferences that are strict partial orders [39]. Two important families of aggregation functions have been thoroughly studied in the last three decades [40]: Choquet integrals [41–44,40] on cardinal scales that generalize the weighted average, and Sugeno integrals [45–48,44] on ordinal scales that generalize the median. Being defined for non-additive measures, these two integrals can take into account possible interactions between evaluation criteria (for instance, there is a synergy between two criteria if the sum of their weights is smaller than the weight of their union, in the case of a Choquet integral). Integral-based aggregations have been also extended to bipolar scales [49–52], encompassing models such as cumulative prospect theory [53].

The use of decision criteria, and more generally aggregation functions, reduces the comparison of alternatives to the comparison of single numbers, which naturally leads to maximization or minimization problems. Thus, aggregation functions provide both a global evaluation of alternatives and a basis for rank-ordering them. As it is well known, Pareto ordering is only defined between dominated alternatives. When comparing two vectors, one may only consider the components for which the values are different, and aggregate these discriminating values, giving rise, for instance, to discrimin or discrimax orderings when min or max are used for the aggregation [54,32]. If all the components have equal importance, then the idea of not taking into account identical values in the comparison may be applied to the vectors once their components have been increasingly or decreasingly reordered, giving rise to leximin and leximax complete preorders [55]. Leximin ordering refines the discrimin ordering, which itself refines both the Pareto ordering and the min-based ordering. Beyond Pareto, other orderings are of interest for comparing vectors of numerical values, such as Lorenz dominance (associated with the Pigou–Dalton transfer principle). They were originally introduced in economics for comparing distributions of incomes [56, 55]; see [57,58] for examples of AI applications.