Arguing About Preferences And Decisions

T.L. van der Weide F. Dignum J.-J. Ch. Meyer H. Prakken
G.A.W. Vreeswijk

Universiteit Utrecht

Abstract

Making a rational decision requires the decision maker's preferences. However, in many situations it is not clear what to prefer. For example, the outcome of a decision may never been encountered before, there may be are very different aspects that matter, or it may be difficult how to weigh short-term and long-term consequences. Furthermore, artificial agents may be required to justify and discuss their decisions to others. Designers must communicate their wishes to artificial agents. Existing research does not address how to reason about what to prefer. Therefore, this paper addresses how to reason about preferences. For this, argumentation is used because it allows to justify and attack preference statements. To be able to justify and question preferences, we propose a qualitative model of preferences that is inspired by multi-attribute utility theory.

Introduction

In many applications where artificial agents are used, agents need to cooperate with either other artificial agents and/or human users. This requires the agent the explain its preferences and decisions and to understand the user's preferences. In this paper, we use as an example that a human user wants to buy a house, but the user does not know what house to prefer. An artificial agent supports the user by arguing about what house is best according to the user's preferences. The problem with buying a house is not that there are too many houses available, but that even deciding between two houses is difficult because so many different aspects matter and therefore it is hard to determine what house to prefer.

In human dialogue, argumentation is typically used to explain decisions and preferences. In recent years, researchers have investigated how argumentation can be used in decision-making. The advantages of using argumentation are that it is intuitive to explain things and that it can handle inconsistent and incomplete information. A possible argumentation framework for this purpose is the one proposed in [1], which enables constructing arguments concerning decisions on the basis of what goals agents pursue. However, it is not possible to discuss what goals an agent has. By understanding better why the user has certain goals, the agent can better determine what is preferred. For example, the agent could also ask the user why he has the goal to live in a house bigger than $60m^2$ and why the user does not look at the volume of the house or whether the house has a garden and how much of the space is usable. Suppose that the user explains that he wants a bigger house because a bigger house is more comfortable. Then the agent could argue that other aspects are also important for how comfortable a house is. The user may not have known or forgotten that these aspects matter, or he may disagree with the agent that these aspects influence comfort.

In the framework proposed by [2], the values that people pursue, e.g. values like fairness, friendship or fun, are used to justify and attack having a goal. People use their values as standards or criteria to guide selection and evaluation of actions. For example, suppose that the user pursues the values of fun and of comfort. The user can now justify his goal to live downtown by arguing that it promotes his values of fun and the goal to live in a bigger house because it promotes the value of comfort. Values are typically very general and subjective. What one person considers to be fair or fun, another does not. What constitutes a specific value like fun, comfort, justice, or health often is disputable and therefore it is also disputable whether a goal or an action promotes a certain value. However, in [2] it is not possible to explain or discuss what constitutes a value and consequently it is also not possible to justify or attack that a goal or action promotes or demotes a value.

Approach

In the paper, we introduce a formal argumentation system based on the ASPIC logic as described in [3] to argue about preferences, goals and values. A model is introduced to decompose an agent's preferences into the perspectives and criteria he cares about. Inference rules are then proposed to infer how an agent prefers outcomes from how outcomes are preferred from the perspectives he cares about. Finally, inference rules are proposed to justify goals based on the decomposition of an agent's preferences.

The notion of a *perspective* is introduced and denotes a criterion from which outcomes of decisions can be compared. Each perspective p is associated with an ordering $\leq_p$ over outcomes. For example, outcome ω_1 may be better than outcome ω_2 from the perspective of costs, worse from the perspective of its centrality, indifferent from the perspective of comfort, and perhaps incomparable from the perspective of fun. Support of decisions is done by arguing what decision's outcome a user should prefer most.

Values are represented as perspectives that people use to make decisions. Because some perspectives are abstract and/or subjective, it may be unclear how to compare outcomes from that perspective. This makes it necessary to specify what that abstract and/or subjective perspective means. For example, suppose that the user said that fun is important to him. The perspective of fun is both abstract and subjective. If the user explains that the centrality of the house influences how fun the house is, then the perspective of fun is specified using the more specific perspective of centrality.

Two binary relations over perspectives are introduced to decompose perspectives: positive and negative influence, denoted $p \uparrow q$ and $p \downarrow q$ respectively with p and q perspectives. If perspective p *positively influences* perspective q, then 'the more preferred an outcome is from perspective p, the more preferred it tends to be from perspective q'. This argument scheme is formalized with the defeasible inference rule $p \uparrow q, \omega_1 \leq_p \omega_2 \Rightarrow \omega_1 \leq_q \omega_2$. Similarly, if perspective p *negatively influences* perspective q, then 'the more preferred an outcome is from perspective p, the *less* preferred it tends to be from perspective q', which is formalized with the defeasible inference rule $p \downarrow q, \omega_1 \leq_p \omega_2 \Rightarrow \omega_2 \leq_q \omega_1$. For example, 'centrality positively influences fun' means that the more central a house is, the more fun it tends to be and 'the costs of a house negatively influences agent A's preferences' means 'the more costs, the less A prefers it'.

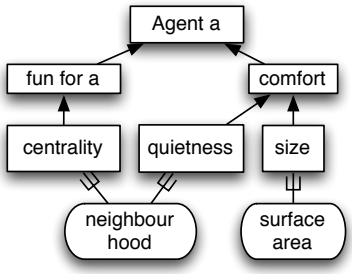

Figure 1: Preference Decomposition

If a perspective is specific, then it can order the attribute-values of an attribute of outcomes. For example, the perspective 'size of house' orders the attribute-values of the attribute 'surface area in m^2'. This is formalized by associating an ordering $\preceq_p$ over the attribute-values of an attribute. The strict inference rule $x(\omega_1) \preceq_p x(\omega_2) \rightarrow \omega_1 \leq_p \omega_2$ is added to the logic to infer how outcomes are ordered from a perspective based on how that perspective orders the outcome's attribute-values. Figure 1 is a sketch of how perspectives can be decomposed into more specific perspectives and finally into attributes of outcomes.

Using the introduced inference rules, the following arguments could be constructed. Arguments A and B both conclude how outcomes ω_1 and ω_2 compare from perspective r but for different reasons. However, argument C conflicts with A and B because of how the attribute x values are preferred from perspective p''.

$$A = \dfrac{q \uparrow r \quad \dfrac{p \downarrow q \quad \omega_2 <_p \omega_1}{\omega_1 <_q \omega_2}}{\omega_1 <_r \omega_2} \qquad B = \dfrac{q \uparrow r \quad \dfrac{p' \uparrow q \quad \omega_1 <_{p'} \omega_2}{\omega_1 <_q \omega_2}}{\omega_1 <_r \omega_2} \qquad C = \dfrac{p'' \uparrow r \quad \dfrac{x(\omega_2) \prec_{p''} x(\omega_1)}{\omega_2 <_{p''} \omega_1}}{\omega_2 <_r \omega_1}$$

Goals are defined as satisfactory sets of attribute-values and are justified using the decomposition of an agent's perspective. For example, agent A has the goal to live downtown because perspective A is influenced by perspective fun, which is influenced by the perspective centrality, which orders the attribute 'neighborhood'. Because the attribute-value 'downtown' is satisfactory, A is justified to have that goal.

References

[1] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413 – 436, 2009.

[2] K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.

[3] M. Caminada and L. Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the National Conference on Artificial Intelligence*, pages 608–613, 2005.