Partial Implication Semantics for Desirable Propositions

Yi Zhou

Department of Computer Science
University of Science and Technology of China
HeFei, 230026 China
zyz@mail.ustc.edu.cn

XiaoPing Chen

Department of Computer Science
University of Science and Technology of China
HeFei, 230026 China
xpchen@ustc.edu.cn

Abstract

Motivational attitudes play an important role in investigations into intelligent agents. One of the key problems of representing and reasoning about motivational attitudes is which propositions are the desirable ones. The answer based on classical logic is that the propositions that logically imply the goal are desirable and the others are not. We argue that this criterion is inadequate for the incomplete knowledge about environments for an agent. In this paper, we present a simple and intuitive semantics---partial implication---for the characterization of desirable propositions. In this semantics, Proposition P is a desirable one with respect to a given goal Q if and only if P is "useful" and "harmless" to Q in any situation. Partial implication is an extension of classical implication. We investigate some fundamental properties of partial implication and discuss some of the potential applications.

Introduction

The study of representing and reasoning about motivational attitudes has attracted a great deal of attention from AI researchers [Bacchus and Grove 1996, Bell and Huang 1997, Chen and Liu 1999, Doyle et al. 1991, Lang et al. 2003, Wellman and Doyle 1991]. It's clear that intelligent agents act under the guidance of their motivational attitudes. In traditional AI planning systems, a goal provides an end state for an agent, and the agent wants to find a sequence of actions to achieve the goal. Similar notions can be found in BDI models [Bratman 1987, Cohen and Levesque 1990], which treats desire as a necessary kind of mental state of agents and formalize it by modal operators. Another approach is QDT [Bacchus and Grove 1996, Boutilier 1994, Doyle et al. 1991]. By using preference among worlds, QDT gives a nice way of formalizing motivational attitudes.

One of the key problems in representing and reasoning about motivational attitudes is the problem of *desirable propositions*. It will be widely accepted that a proposition P is desirable if and only if P is "useful" to the agent. It should be emphasized that desirable propositions are different from desires of agents, since that a proposition P is "useful" to an agent doesn't mean that the agent wants to achieve P. Given a proposition P, when can one say that P is desirable to the agent?

This problem is important. One reason is that an agent's rational actions must be desirable (or motivational), and apparently, desirable propositions are highly related to desirable actions, as we will show later. Desirable propositions are a bridge between the agent's desires and alternative actions. Another reason is that we often decompose our complicated goals into smaller "pieces" in order to reduce the complexity of the problem. Of course, these "pieces" must be desirable. We'll make the concept of "pieces" clear later.

According to the classical propositional logic, a proposition P is desirable if and only if the agent believes that P implies the agent's goal. However, this criterion is inadequate for realistic situations for some reasons. Firstly, the knowledge of an agent is incomplete, and maybe there exists a way to achieve the goal but the agent can't find it out due to her incomplete knowledge. Secondly, even if an agent has complete knowledge, other kinds of resource boundedness will make the agent not to achieve the goal completely. If an action is useful to the agent's goal but cannot completely achieve it, is this action an alternative, or desirable, one? For example, suppose that Alice has a goal "*having some milk and bread*", but milk and bread have all been used out in the home. An idea that comes up is "*having some from the store*". Alice knows that there is bread in the store, but Alice doesn't know whether there is milk to be sold or not there. In classical planning approach, Alice will do nothing because there is no way to completely achieve the goal according to her current knowledge. This is not appropriate or adequate for an autonomous agent in unpredictable environments.

Previous work on qualitative decision theory mainly focus on the preference among alternative propositions. In preference semantics one can determine whether one proposition is more desirable than another and thus make decisions correspondingly. We believe that desirability is of the same importance as preference in rational decision making. There are some significant differences between the desirable ones and undesirable ones. As in the milk case, "*having some milk*" is the desirable one yet "*having some tea*" is not. Although we can represent this distinction in QDT, we cannot explain why the former is desirable and the latter is not, all we know is that the former is more desirable than the latter according to our preference over worlds. In this paper we'll provide an alternative approach to clarify the inherent propositional relation between the agent's goals and the desirable propositions.

We believe that there are different sorts of desirable propositions. In this paper, we concentrate on formalizing a certain kind, called *partial desirable propositions*. We extend the classical implication into partial implication, denoted by $\Gamma \models P \succ Q$, meaning that under any circumstance Γ, if P is true, then Q will be "partially" true. Technically, we employ and generalize the semantic theory L_{mp4c} [Chen and Liu 1999] and prime implicant [Marquis 2000] to formalize partial implication. Unlike L_{mp4c}, here we use a two-valued semantics instead of four-valued one.

The rest of this paper is organized as follows. We clarify the notion of desirable propositions and partial implication in section 2. In section 3 we formalize our criterion by prime implicant and investigate its main properties. Then we define a kind of desirable proposition based on partial implication and show how this definition relates to some key problems in desire representation in section 4. At last we draw our conclusions.

Informal discussion

The concept of desirable propositions which we try to capture in this paper is different from that of *desires* appeared in existing literature. According to existing literature a proposition P is a desire if the agent wants to achieve P. On the other hand, we call a proposition P a *desirable proposition* if P is desirable to some extent, i.e., at least "useful" to the agent's goal. This idea was first stated by Newell [Newell 1982]. In the milk case, does Alice want "*to have some bread and milk from the store*"? Not absolutely. Maybe the distance is so far that Alice doesn't want to go to the store. What Alice wants is just "*having milk and bread*" at the right moment, "*having the bread and milk from store*" is only useful with respect to the goal. Before she decides to go to the store, this proposition is not her traditional desires, that is, she wants "*to have bread and milk from store*" only after she makes the decision.

For the second, a desirable proposition P is relative to one or some of the traditional desires. In the previous example, "*having the bread and milk from store*" is a desirable proposition with respect to "*having some bread and milk*". Suppose that Alice has another goal "*not to be tired*", then "*having milk and bread from store*" is not an "absolutely desirable proposition" since Alice suffers from a long walk to go to the store.

And another important thing is that the agent's desirable propositions are relative to the agent's beliefs. Suppose that someone tells Alice that the milk in the store has been sold out and Alice believes it, then she would not take "*having some bread and milk from store*" as her desirable proposition.

Ideas relevant to desirable propositions or desirable actions can be found in AI literature elsewhere, such as the famous means-end analysis, Γ in BOID architecture [Broersen et al. 2001], and so on. Apparently, generating options and weighting of options are both necessary for a rational agent. The goal of AI planning systems is to generate action sequences to achieve given goals from initial state. If either of the action sequences does not completely achieve the goal, there is no difference between them in traditional AI planning systems. Haddawy and Hanks [Haddawy and Hanks 1998] criticized it and used a function from propositions to a real number to represent the degree of satisfaction of goals. In this paper, we provide partial implication semantics for desirable propositions from a symbolic point of view. We focus on the partial implication relationships between desirable propositions and goals but not the degree of satisfaction between them. Partial implication is an extension of classical implication which means that under a circumstance Γ, if P is true, then Q will be "partially" true.

Partial implication is closed under "usefulness" while classical implication is closed under deduction. Intuitively, that P partially implies Q can be understood as that P is "useful and harmless" to Q in any situation, where *P's* being useful to Q means that the realization of P will cause partial realization of Q, and *P's* being harmless to Q means that the realization of P will not damage partial realization of Q. We will give formal definitions of usefulness and harmlessness in our semantics.

Let’s recall the milk case. Let x= *having some milk*, y= *having some bread*, then the goal of Alice can be represented as $G(x \wedge y)$. Are these following propositions---x, z, $x \wedge z$, $x \wedge \neg y$, $x \vee y$, $x \vee z$ desirable propositions of Alice (z denotes *having some tea*)? According to the discussion above, the conclusion is that x, $x \wedge z$, $x \vee y$ are desirable and the others are not. The reason is that $x \wedge \neg y$ is useful but also harmful to $x \wedge y$; since $\neg x \wedge z$ implies $x \vee z$, $x \vee z$ is not useful to $x \wedge y$ “in any situation”. However, in some realistic situation, we treat $x \vee z$ as a desirable proposition with respect to $x \wedge y$, we think that this is another kind of desirable proposition.

Partial implication semantics

We will restrict our discussion within a propositional language, denoted by L. The formulas of L are formed as usual from a set of atoms, $Atom$= $\{x_1, x_2 \ldots\}$, and the standard connectives $\neg, \vee, \wedge, \rightarrow$ and $\leftrightarrow$. For any subset A of $Atom$, let $L(A) = A \cup \{\neg a \mid a \in A\}$ be the set of literals composed from A. Let Γ, Γ', etc denote sets of consistent propositional formulas in L; x, y atoms; l literals; and P, Q, R formulas in L. $Atom(P)$ and $Atom(\Gamma)$ denote the set of atoms appeared in P and Γ respectively.

Definition 1: (Γ-implicant) A consistent conjunction of literals π is a Γ-implicant of formula P, if:

(1) π is consistent with Γ,

(2) π and Γ satisfies P.

We write $\Gamma[P]$ to denote the set of the Γ-implicants of P. A Γ-implicant is a partial truth-value assignment, under which some atoms may have no truth-value assigned. It represents incomplete knowledge or recognition state of an agent. But one should notice that sometimes complete knowledge is unnecessary and the truth-value of a formula is determined even under a Γ-implicant. For example, let Γ be empty, then a Γ-implicant of formula $x \vee y$ is $\{x\}$, only x is assigned truth, all the others are not assigned. In this case, the truth-value of $x \vee y$ does not depend on that of y. Thus a Γ-implicant can be understood as a reduced representation of the relevant state as well.

Definition 2 (Γ-prime implicant) A consistent conjunction of literals π is a Γ-prime implicant of formula P if:

(1) π is a Γ-implicant of P,

(2) There isn’t another Γ-implicant π' of P such that $\pi' \subset \pi$.

We write $\Gamma(P)$ to denote the set of the Γ- prime implicant of P. We have that $\Gamma(P) \subseteq \Gamma[P]$. If $Atom(P) \cup Atom(\Gamma)$ is finite, then $\Gamma(P)$ is finite. Intuitively, a Γ-implicant of P means a way to achieve P under Γ, but maybe redundant, and yet, a Γ-prime implicant of P means an exact way to achieve P, all atoms in the prime implicant are necessary. Suppose that a Γ-implicant is a way to achieve formulas, and then a Γ-prime implicant is a “least” way to achieve formulas. Each atom in Γ-prime implicant is an “essential part”. If another proposition makes some of the literal in a Γ-prime implicant of P true, then it is useful to P. Moreover, if the proposition doesn’t make the rest literals false, then it is harmless to P. A subset of a Γ-prime implicant of P is what we call a “pieces” of P, which is simpler than the original goal P. In particular, if Γ is empty, we say that π is a prime implicant of P.

Let $-\pi$ be the set of negations of the literals in π. Similar to L_{mp4c}, we define partial implication based on prime implicant.

Definition 3 (Partial implication) We say that P partially implies Q under Γ, denoted by $\Gamma \models P \succ Q$, if:

(1) $\Gamma(P)$ is not empty.

(2) For each $\pi \in \Gamma(P)$, there exists $\pi' \in \Gamma(Q)$, such that $\pi \cap \pi'$ is not empty and $\pi \cap -\pi'$ is empty.

The Γ- prime implicant of P π plays a dual role. Firstly, it captures the “any situation”, all the situations must be consistent to Γ and P will be true in the situation. And the set $\Gamma(P)$ means all possible ways to achieve P; Secondly, π is useful (achieves part of Q, $\pi \cap \pi'$ is not empty) and harmless (we can achieve Q based on the contribution of P, $\pi \cap -\pi'$ is empty) to Q, and then, P is useful and harmless to Q through π and π'. Thus we formally define our intuitive definition of partial implication by using prime implicant.

In partial implication semantics, principle of substitution doesn’t hold. For example, we can get $\models x \succ x \wedge y$ when Γ is empty, but it is not the case that $\models P \succ P \wedge Q$ for all formulas P and Q (let P= $x \vee y$ and Q=$\neg y$). We enumerate some instances of the partial implication to illustrate its characteristics over atoms first. As a matter of convenience, we omit Γ when it is empty.

Proposition 1 (Partial implication relationships over atoms)

(p-1) $\models x \succ x \wedge y$;

(p-2) $\models x \succ x \vee y$;

(p-3) $\models x \wedge y \succ x$;

(p-4) $\not\models x \vee y \succ x$;

(p-5) $\not\models y \succ x$;

(p-6) $\not\models x \wedge \neg y \succ x \wedge y$;

(p-7) $\models x \wedge z \succ x \wedge y$;

(p-8) $\models x \vee y \succ (x \vee y) \wedge z$;

(p-9) $\models x \succ x \wedge (\neg x \vee y)$;

(p-10) $\not\models \neg x \succ x \wedge (\neg x \vee y)$;

(p-11) $\models x \succ (x \wedge y) \vee (\neg x \wedge \neg y)$;

(p-12) $\models \neg x \succ (x \wedge y) \vee (\neg x \wedge \neg y)$;

(p-13) $\{y \rightarrow x\} \models y \succ x$;

(p-14) $\{\neg y\} \models x \vee y \succ x$;
(p-15) $\{\neg y\} \not\models x \wedge y \succ x$;
(p-16) $\{z \rightarrow y, y \rightarrow x\} \models z \succ x$.

From (p-1) to (p-16), one can see that partial implication is different from classical propositional implication. Classical propositional logic is suitable for reasoning about beliefs, not suitable for desires. In contrast with (p-5), (p-13) illustrates the uses of preconditions. Comparison of (p1-11) and (p1-12) indicates that a condition and its negation can partially imply the same proposition in our model.

Theorem 2 (Equivalence) If $\Gamma \models P \leftrightarrow Q$, then $\Gamma \models P \succ R$ implies $\Gamma \models Q \succ R$; $\Gamma \models R \succ P$ implies $\Gamma \models R \succ Q$.

Theorem 3 (Relationship to classical implication) If $\Gamma \models P \rightarrow Q$ and both P and Q are non-trivial under Γ, then $\Gamma \models P \succ Q$.

From theorem 3 it follows that partial implication is an extension of classical implication.

Theorem 4 (Non-triviality) If P or Q is trivial under Γ, then that $\Gamma \models P \succ Q$ doesn't hold.

This theorem illustrates that every trivial formula does not partially imply other formulas and can not be partially implied either.

Lemma 5 Suppose Γ is empty, if $Atom(P) \cap Atom(Q)$ is empty, then $\Gamma(P \vee Q) = \Gamma(P) \cup \Gamma(Q)$, $\Gamma(P \wedge Q) = \{\pi \cup \pi' \mid \pi \in \Gamma(P), \pi' \in \Gamma(Q)\}$.

Theorem 6 (Conjunctive decomposition) If $Atom(P) \cap Atom(Q)$ is empty, and P, Q are non-trivial, then $\models P \succ P \wedge Q$.

Theorem 7 (Disjunctive decomposition) If $Atom(P) \cap Atom(Q)$ is empty, and P, Q are non-trivial, then $\models P \succ P \vee Q$.

Proposition 8 (Relevant) If $\models P \succ Q$, then $Atom(P) \cap Atom(Q)$ is not empty.

Unlike relevance logic, partial implication focuses on "partial" relationships between formulas. The antecedent needn't imply the consequent. For example, (p-1) and (p-7) in proposition 1 show that some partial implication relationship don't hold in relevance logic.

Property 9 (Non-monotonic) There exists a formula P and a Γ-implicant π of P such that π is not a Γ'-implicant of P, where $\Gamma' = \Gamma \cup \{Q\}$.

An example is that Γ is empty, $P = x \wedge y$, $Q = x$, $\pi = \{x, y\}$. This property is also true for Γ- prime implicant. So non-monotonicity can be found everywhere in partial implication semantics. Such as the non-monotonicity of Γ-implicant, that of Γ- prime implicant, precondition set Γ, etc. Among them, property 9 is the most fundamental one, all the other non-monotonicity of partial implication can be derived from it. Non-monotonicity is an inherent property of desirable propositions. It's not the same as the non-monotonicity of beliefs.

Property 10 (Decomposition) If any intersection of sets $Atom(P_1)$, $Atom(P_2)$ ... $Atom(P_n)$ is empty, we can replace these formulas by atoms.

This means that we can reduce the complexity in partial implication semantics due to the "independence" knowledge of agent. Similar notions can be found in QDT [Bacchus and Grove 1996]. Theorem 6 and theorem 7 are special examples.

Property 11 (Non-transitivity) $\Gamma \models P \succ Q$ and $\Gamma \models Q \succ R$ doesn't imply $\Gamma \models P \succ R$.

The reason of the untenable of transitivity is that in some situation, the "usefulness" of P with respect to Q is just irrelevant and redundant for the "usefulness" of Q with respect to R. As an example, suppose Γ is empty, then $\models x \succ x \wedge y$ and $\models x \wedge y \succ y$ hold, but $\models x \succ y$ does not hold. We will elsewhere strengthen partial implication into strong one which satisfies transitivity.

Desirable proposition and related works

Desires themselves are all desirable propositions. Perhaps the simplest definition of desirable propositions could be given this way: propositions that imply (achieve) the agent's goal under the agent's beliefs are the desirable ones. We call these propositions (actions) strict desirable propositions (actions).

Definition 4 (Strict desirable propositions) Let Q is the agent's desire and Γ is the agent's belief set. If $\Gamma \models P \rightarrow Q$, then the P is called a strict desirable proposition of the agent *wrt* Q.

We have pointed out in section 1 that this criterion is not the unique one, and now we define partial desirable propositions based on the semantics of partial implication.

Definition 5 (Partial desirable propositions) Let Q is the agent's desire and Γ is the agent's belief set. If $\Gamma \models P \succ Q$, then the P is called a partial desirable proposition of the agent *wrt* Q.

The differences between partial desirable propositions and strict desire propositions have been discussed in section 3. From theorem 4, it follows that all nontrivial cases of the strict desirable propositions is included in partial desirable propositions. In fact, the only inadequacy of strict desirable propositions lies in that it does not cover all of the interesting cases. Some of these can be found in Proposition 1.

Let’s go back to the example given in section 1. Alice has a goal “*having some milk and bread*”, but milk and bread have all been used out in the home. Alice knows that there is bread to sell in the store, but Alice doesn’t know whether there is milk to be sold. This is a typical incomplete knowledge in planning. In the classical planning approach, Alice will do nothing because there isn’t a complete way to achieve the goal due to current knowledge. Under the criterion of partial desirable proposition, “*to go to store*” is a rational alternative action because *having some bread* is a partial desirable proposition of Alice.

In this example there is no plan to completely achieve the goal due to the incomplete knowledge of the agent; in that case, we can take partial desirable propositions as an alternative. It is true for many realistic situations.

The study of desirability is also related to the well-known BDI modeling. A relevant question is where do intentions come from? Cohen and Levesque gave an answer that intentions are choices with commitment to goals [Cohen and Levesque 1990]. However, we believe that intentions should be chosen from desirable ones, but not from goals directly. The distinction is significant in that one should derive desirable propositions from goals and choose some of them with commitment as one’s intentions. This provides a possibility of developing some mechanism for deriving intentions automatically and thus a chance of connecting BDI models more directly with certain types of behaviors.

An essential problem in representation of desires is that an agent’s desires seem to relate with her beliefs. Such as the example of rain and umbrella [Wellman and Doyle 1991], suppose there are three independent logical propositions: W, “I’m wet”; R, “It rains”; U, “I’m going out with an umbrella”. We assume that $\neg U$ is preferred to U, but U is preferred to $\neg U$ by given R. Boutilier [Boutilier 1994] formalized it by conditional desires, which $I(B|A)$ means that “If A, do B”. The example can be represented as $\{I(\neg U|True), I(U|R)\}$. We here elaborate the examples by desirable propositions. The mental states of the agent can be represented as $\{D(\neg W)$, $B(\neg R \rightarrow \neg W)$, $B(R \wedge U \rightarrow \neg W)\}$. We can derive from partial implication semantics that $\{R, \neg R \rightarrow \neg W, R \wedge U \rightarrow \neg W\} \models U \succ \neg W$; $\{\neg R, \neg R \rightarrow \neg W, R \wedge U \rightarrow \neg W\} \not\models U \succ \neg W$. So U partially implies $\neg W$ by given R and does not by given $\neg R$, U is a desirable proposition by given R and is not by given $\neg R$. The studies of desirable propositions provide another perspective on the relation between desires and beliefs.

Conclusion

This paper contains two main contributions. First, we point out the difference between traditional “desires” and “desirable propositions”, and analyze the roles of desirable propositions. We expound the inadequacy of the simple criterion of desirable propositions which is based on classical logic. Second, we present a logical semantics to capture the notion of partial implication---being both useful and harmless. Based on this semantics we define a novel criterion of desirable propositions. Our analyses show that the new criterion is necessary and appropriate in some realistic situations.

Notice that strict and partial desirable propositions are not the only possible criterions of desirable propositions, so further investigations into desirable propositions might be taken in the future. Another deserving work is to integrate desirability and preference. Also the first-order formalization of partial implication is worth pursuing.

Acknowledgement

This work has been supported by NSFC of China (60275024). We thank Zhisheng Huang for his valuable suggestions and comments, and also thank Yajun Wang , Zhuo Huang and ShiZhuo Zhu for their useful comments. Specially, we thank the anonymous reviewers of this paper for their advices.

References

[1] F. Bacchus and A. Grove, 1996. Utility independence in qualitative decision theory. In proceedings of KR’96, pages 542-552. Morgan Kaufmann, San Francisco, California.

[2] J. Bell and Z. Huang. 1997. Dynamic goal hierarchies. In J. Doyle and R. H. Thomason, editors, Working Papers of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning, pages 9-17, Menlo Park, California. AAAI.

[3] C. Boutilier. 1994. Toward a logic for qualitative decision theory. In Proceedings of KR’94, 75-86. Morgan Kaufmann.

[4] M. E. Bratman. 1987. Intentions, Plans, and practical reasoning. Harvard University Press, Cambridge Mass.

[5] J. Broersen, M. Dastani, Z. Huang, J. Hulstijin and van der Torre. 2001. The BOID architecture. In proceedings of AGENT’01.

[6] Xiaoping Chen and Guiquan Liu. 1999. A logic of intention, in: Proceedings of IJCAI-99, 1999, 172-177.

[7] P. R. Cohen, H. J. Levesque. 1990. Intention is Choice With Commitment, *Artificial Intelligence*, 42, 213-261.

[8] F. Dignum, D. Kinny and L. Sonenberg, 1997. Formalizing motivational attitudes of agents: On desires, obligations and norms. In proceedings of the 2nd IWCEE on MAS.

[9] J. Doyle, Y, Shoham, and M. P. Wellman. 1991. A logic of relative desire, in: Proc. of the 6th IS on Methodologies for Intelligent Systems, Pages 16-31.

[10] P. Haddawy and S. Hanks. 1998. Utility models for goal-directed decision-theoretic planners. Computational Intelligence, 14(3): 392-429.

[11] J. Lang, L. van der Torre, E. Weydert. 2003. Hidden Uncertainty in the Logical Representation of Desires. In Proceedings of IJCAI'03.

[12] P. Marquis. 2000. Consequence finding algorithms. In D.Gabby, *Handbook of Defeasible reasoning and Uncertainty Management Systems*(V), pp, 41-145. Kluwer.

[13] A. Newell. 1982. The knowledge level. Artificial Intelligence, 18(1): 87-127.

[14] M. P. Wellman and J. Doyle. Preferential semantics for goals. 1991. In proceedings of AAAI'91, pages 698-703, Menlo Park, California. AAAI Press.