基于关系的两维意向结构*

李 斌[1,2+], 朱朝晖[1], 陈韬略[1], 吕 建[1], 朱梧槚[1]

[1](南京大学 计算机软件新技术国家重点实验室,江苏 南京 210093)

[2](扬州大学 计算机科学与技术系,江苏 扬州 225009)

Two-Dimensional Intention Structure Based on Relation Structure

LI Bin[1,2+], ZHU Zhao-Hui[1], CHEN Tao-Lue[1], LÜ Jian[1], ZHU Wu-Jia[1]

[1](State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

[2](Department of Computer Science and Technology, Yangzhou University, Yangzhou 225009, China)

+ Corresponding author: Phn: +86-25-3303169, E-mail: libin@softlab.nju.edu.cn, http://www.nju.edu.cn

Received 2003-01-10; Accepted 2003-11-10

Li B, Zhu ZH, Chen TL, Lü J, Zhu WJ. Two-Dimensional intention structure based on relation structure. *Journal of Software*, 2004,15(4):512~521.

http://www.jos.org.cn/1000-9825/15/512.htm

Abstract: This paper aims to introduce a new representation framework of cognitive state of agents, which includes depictions for intention, belief, and goal, and supplies a necessary theoretical basis for constructing agent architecture. In this formalization, all intentions for a goal form a structure with two orders, one represents the temporal order and the other represents the relevant relations among intentions. At the same time, this paper investigates the relationships among beliefs, intentions, and goals. Since this framework gets rid of modal operators to depict intention, intention base, temporal and relevant relation are be adopted to represent intention when agents are constructed, which can narrow the gap between the agent model and agent architecture.

Key words: agent theory; agent architecture; BDI model; intention theory; relation structure

摘 要: 从建构 agent 角度出发,提出了一个基于关系结构的包括 agent 意向、信念以及目标等认知状态的框架.在此框架中,实现目标的意向形成了两维序结构,其中一维表示意向间的时序关系,另一维表示意向间的相干

* Supported by the National Natural Science Foundation of China under Grant Nos.60273034, 60233010 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant Nos.2002AA116010, 2001AA113110 (国家高技术研究发展计划(863)); the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312002 (国家重点基础研究发展规划(973)); the Natural Science Foundation and High Technology Project of Jiangsu Province of China under Grant Nos.BK2002203, BK2002409, BG2001012 (江苏省自然科学基金和高技术项目); the Key Project of Natural Science Foundation of the Education Department of Jiangsu Province of China under Grant No.01KJB520010 (江苏省教育厅自然科学基金重点项目)

LI Bin was born in 1965. He is a postdoctor at Nanjing University. His research areas are agent technology and artificial intelligence. **ZHU Zhao-Hui** was born in 1970. He is an associate professor at Nanjing University of Aeronautics & Astronautics. His research area is artificial intelligence. **CHEN Tao-Lue** was born in 1980. His research area is agent technology. **LÜ Jian** was born in 1960. He is a professor and doctoral supervisor at Nanjing University. His research areas are agent technology, object-oriented technology. **ZHU Wu-Jia** was born in 1955. He is a professor and doctoral supervisor at Nanjing University of Aeronautics & Astronautics. His research area is artificial intelligence.

关系,在此基础上,研究了信念、意向和目标的相互关系.因为摒弃了传统的用模态算子来刻画 agent 的意向的方法,所以在构建 agent 时,可以直接采用意向库以及意向间的时序、相干关系来表示 agent 的意向,从而缩小了 agent 理论模型与实际 agent 结构之间的差异,为 agent 结构的建立提供了必要的理论基础.

关键词: agent 理论;agent 结构;BDI 模型;意向理论;关系结构

中图法分类号: TP18 **文献标识码:** A

One of the most important tasks for agent research is the design and construction of agents to satisfy the properties specified by agent theory[1]. BDI model may be the most popular agent theory model and has been well studied in the AI literature[2], thus it has given rise to be BDI architectures where the elements of belief bases, goal bases, intention bases, and plan libraries are central. The main tasks of research on BDI model are to bring the relationships among beliefs, goals and intentions out, and several formalizations of cognitive state including intentions and beliefs have appeared in the recent literature[3~8]. Since normal modal logics (NMLs) have proven to be a useful tool in modeling the cognitive attitudes of belief and knowledge, many researchers have employed NMLs to model intentions. However, since intention and belief are different, NMLs are ill-suited for modeling intentions. Furthermore, although the BDI model has been applied quite successfully, an as yet ongoing frustration among agent researchers is the gap between the formal (BDI) model and the (BDI-based) architectures in the sense that one would like to use the former to specify the latter formally and prove the formal properties about these. But this has not been shown to be possible, as the 'distance' between the two 'worlds' is too large. Konolige and Pollack[5] presented an alternative model of intentions based on *possible worlds* and *cognitive structures*. Since K&P's framework is based simply on set-theoretic operations on *possible worlds*, it still has some important limitations[6]. Furthermore, because of the concentration on the semantics, K&P's work provides neither the descriptions of the relations between beliefs and intentions nor those of relations among intentions in the syntactical categories. Thus, it is difficult to establish a dynamic intention theory on the basis of their framework.

From the point of view of agent construction, this paper presents a new formalization of cognitive state including intentions, beliefs, and goals based on relation structure. In this framework, all intentions for a goal form a structure with two orders which represent the temporal order and relative relationship among intentions respectively. This structural intention model allows us to inquire into something that has not been explored extensively in the previous work, such as the change and generation of intentions. Furthermore, our framework supplies a necessary theoretical basis for constructing agent architecture. The reason for that is our framework gets rid of the modal operators to depict intention, so we can adopt intention base, temporal, and relevant relation to represent intention when we construct agent, which can narrow the gap between agent model and agent architecture.

In Section 1, we expound the intuitive idea of this paper. Section 2 constitutes the technical core of our paper: there we develop our formalization. In Section 3, we end the paper with some conclusions.

1 Background and Motivation

In this section, we will discuss some characteristics of intentions and explain the intuitive idea of this paper. We begin with comparing beliefs with intentions.

· **Non-Pure-Logical relationship** One of the important differences between intentions and beliefs lies in the fact that the relations among beliefs are logical and the relations among intentions are not. According to our intuition, the logical consequences of basic beliefs are also rational beliefs. Thus, given a set of basic beliefs, we may determine a maximal belief set by means of logical inference. It is why logical omniscience may be assumed as an idealization. This assumption only contradicts agent's limited capability of computing, and doesn't lead to results violating our intuition. However, consequential closure cannot be assumed for intentions, even as an idealization. It is well known that there exists side-effect problem[6] in the formalizations of intentions based on NMLs. This problem is more serious and harmful to intention than belief: logically consequential closure cannot be assumed for intention, even as an idealization; not all the consequences of an agent's intention are intentions of the agent, even the consequences he/she has anticipated. The similar problem, so-called logical omniscience, also appears in the researching on belief and knowledge, many researchers have studied this problem. However, the later justly idealizes the reasoning capability of agent, and does not induce conclusions violating our intuitions. Although several strategies have been presented to make the logic *side-effect* free, it is difficult to eliminate the following so-called *equivalent side-effect problem* in (normal or nonnormal) modal logics[6], which is a special form of the *side-effect problem*: If $\alpha\leftrightarrow\beta$ then $INT(\alpha)\leftrightarrow INT(\beta)$. From the point of view of bounded rationality and resource-boundedness, *equivalent side-effect* is inappropriate and harmful: logical equivalencies are not "cognitive equivalencies".

In fact, the *side-effect problem* reflects that modal frameworks are inappropriate tools used to formalize intentions. The relations among intentions are not pure logical, and the generation and change of intentions are restrained by agent's commonsense reasoning and also dependent on agent's choices and strategies to a great extent. In other words, given goals and beliefs, agent's reasoning, choices, and strategies jointly determine his intentions.

· **Temporal order among intentions** Beliefs mirror agent's views on the world at a moment, the whole beliefs of agent may be simply treated as a set satisfying certain logical conditions. However, the similar treatment of intentions is not enough to capture their characteristics. Since intentions reflect agent's decisions about how to achieve his goal, representing the temporal order among intentions is essential for modeling intentions. Otherwise, it may lead to irrational results. For instance, the following axioms are presented in Ref.[6]:

(1) *Conjunctive composition*: $\models INT(\alpha)\wedge INT(\beta)\supset INT(\alpha\wedge\beta)$,

(2) *Disjunctive closure*: $\models INT(\alpha\vee\beta)\wedge INT(\neg\alpha)\supset INT(\beta)$.

Example 1. As an example, consider the simple blocks world which consists of three blocks A, B and C. The initial state of the blocks world is that block A and block B are on the table, block C is on top of block A. An agent intends to change the world from the initial state to the target state that is a tower in the order C, A, and B from top to bottom.

Suppose that the agent can move one block at a time and has three intentions: $INT(\neg on(C,A))$, $INT(on(A,B))$ and $INT(on(C,A))$, where on is a binary predicate, $on(x,y)$ means that x is on y. From the background, we know there exists a temporal order among the above intentions, that is, agent intends to achieve the state $\neg on(C,A)$ firstly, $on(A,B)$ secondly and $on(C,A)$ finally. Applying the axiom *Conjunctive composition* simply without considering the temporal order will lead to the irrational intention $INT(\neg on(C,A)\wedge on(C,A))$.

In our opinions, since there is a temporal order among intentions, the axioms of the simple logical combinations among intentions, such as the axioms *Conjunctive composition* and *Disjunctive closure*, should be turned down.

In this paper, we introduce an order (denoted by $\prec_t$) defined on the set of intentions, which not only represents the temporal order among intentions but also plays an important role in dealing with the relations between beliefs and intentions.

It is well known that one of the important tasks in intention theory is to model the interactions between intentions and beliefs. In the literature, many properties have been presented and many authors have noticed that temporal order is necessary for the researching in this aspect, and some temporal frameworks, such as linear time logic and branching time logic[8], have been adopted. Since those temporal frameworks only provide a rough depiction on temporal order, their limited expressive power is not enough in dealing with the generation and change of intentions. The generation of intentions depends on the beliefs about not only the current world but also the future. This may be described as follows: agent generates the intention i_1 on the basis of the current belief set B_0 and his goal g, and predicts the situation of the world in which i_1 is achieved and then generates the corresponding belief set B_1 concerning this world by updating B_0. Furthermore, he generates intention i_2 based on B_1, and so on.

The process described above reveals that the generation of intentions always depends on agent's prediction concerning the results caused by the realization of some pre-existing intentions. In coming paragraphs, we will introduce the operator *pred* to represent agent's predicting capacity. Furthermore, in the static intention framework of this paper, we will give some rational constraints on the relations between intentions and a sequence of belief sets obtained by means of the operator *pred* and the order $\prec_t$. This static framework is appropriate for investigating the change of intentions.

· **Consistency and compatibility** It is well known that agent's belief set must be consistent. Some authors have also presented some consistency axioms for intentions. The following axioms appear in Ref.[6]:

(1) *Intention consistency I*: $\models\neg INT(\alpha\wedge\neg\alpha)$,

(2) *Intention consistency II*: $\models INT(\alpha)\supset\neg INT(\neg\alpha)$.

In the above two axioms, the axiom *Intention consistency I* is rational, which reflects the fact that contradictory objects cannot serve as intentions for a rational agent. However, from our points of view, the axiom *Intention consistency II* is irrational. For instance, in the Example 1, the agent has the intentions $INT(\neg on(C,A))$ and $INT(on(C,A))$ simultaneously and plans to achieve them respectively in a proper time.

As a matter of fact, we think that the following principle, namely the Principle of *Compatibility*, is more appropriate than the consistency axiom for intentions.

The Principle of Compatibility: For each intention α, agent is supposed to believe that the realization of the intentions being scheduled to be achieved before a, doesn't make the realization of α impossible.

Example 2. Consider the following scenario. There exist two birds b_1, b_2 in a tree, an agent has a shotgun. In the light of *Compatibility Principle*, $shot(b_1)$ and $shot(b_2)$ don't coexist as intentions for the agent who believes that shooting one bird may cause the other to fly away. Thus, the agent may hold one at most as his intention between $shot(b_1)$ and $shot(b_2)$, whereas the axiom *Intention consistency I* and *II* could not exclude the possibility of holding $shot(b_1)$ and $shot(b_2)$ as intentions simultaneously.

The following property is valid in K&P's framework[5]:

Epistemic consistency: $I(\alpha)\supset\neg B(\neg\alpha)$

Recall that the B operator in the above represents all futures that the agent believes might occur, hence the above formula means that agent should believe that each of his intentions is achievable. It is easy to see that *Epistemic consistency* and *Compatibility Principle* have something in common, that is to say, they all confirm that agent should believe in the achievability of his intentions. But there exists a fundamental difference between them, that is, the axiom *Epistemic consistency* concentrates on each intention individually, whereas *Compatibility Principle* focuses on all intentions as a whole and takes account of the interaction of intentions.

· **Hierarchy and evolution** We hold that any intentions of a rational agent should serve the realization of his goal, that means whatever a rational agent does is for certain purposes. To achieve his goal, agent has a plan which is a sequence of actions and which determines how to achieve the goal. A plan is only a part of agent's intentions, and is the final outcome of making the relatively rough intentions. The evolution of intentions brings about the relative hierarchies among intentions, the upper intention is rougher than the lower intention, and the goal is the roughest intention in the corresponding intention set in which all intentions serve this goal.

It is necessary to point out that the similar idea also appears in Ref.[5], where K&P introduce the concept *graph fragment* to represent the relativity among intentions. In this paper, we introduce an order on the set of intentions (denoted by $\prec_r$) to depict the evolutionary tendency of intentions.

2 Two-Dimensional Structure of Intention

In this section, we will present a formal description of the *intention structure*. We begin with an example to explain the idea about the intention mentioned above.

2.1 Adam in a blocks world

Consider one agent, Adam, living in a blocks world containing three blocks A, B, and C. We use *clear*(X) to describe that no block is on top of X, the meaning of *on*(X,Y) is the same as above. We allow only one atomic action: *put*(X,Y), which has the effect of Y being placed on X. Adam has the ability of performing a *put*(X,Y) action if and only if X and Y are clear and Y is not identical to X. Adam commits to implement a goal, i.e., Adam intends to change the world from the initial state to the target state depicted in Example 1. So Adam's initial belief is B_0={*on*(C,A), *clear*(B),*clear*(C)}, initial goal is G_0=*on*(A,B)∧*on*(C,A)∧*clear* (C), and level 0 intention I_0(represented as P_0 in Fig.1) is G_0. To implement I_0, First, Adam generates some rough abstract plans $I_{11}, I_{12}, \ldots, I_{1n}$, which can attain intention I_0. Second, according to some choice mechanism, Adam commits a plan I_{1i} as his level 1 intention. For example, I_{1i}={P_{11},P_{12}}, P_{11}=*on*(A,B)∧*clear*(A)∧ *clear*(C), P_{12}=*put*(C,A). So, level 0 intention P_0 is refined into two level 1 intentions P_{11} and P_{12}, depicted as Fig.1. It is notable that there are some relations among intentions P_0, P_{11}, and P_{12}. Intention P_{11} must be achieved before intention P_{12}, i.e. there is a temporal order relation between intentions P_{11} and P_{12}. The achievement of intention P_{11} and intention P_{12} can bring about the achievement of intention P_0, so there are relevant relations among intentions P_0, P_{11} and P_{12}. Furthermore, intention P_{11} can be refined into intention P_{21}=*clear*(A)∧*clear*(B)∧*clear*(C) and intention P_{22}=*put*(A,B), intention P_{21} can be refined into intention P_{31}=*put*(C,*Table*), and so on. The refinement graph is depicted in Fig.1.

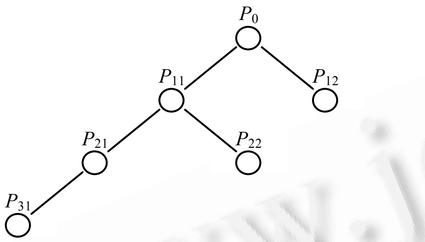

Fig.1 Intention structure refinement graph

In conclusion, we think that there are relations among all intentions for a goal, one is the temporal order relation and the other is the relevant relation, i.e. all intentions for a goal form a structure with two orders. Before presenting a formal description of *intention structure*, we begin with some concepts.

2.2 Basic concepts

As mentioned above, in this paper, all intentions serving a given goal g form a finite structure which is called *intention structure for the goal g*. An intention itself may be a well formed formula(*wff*) or an action expression. In the following, the set of *wffs*(or, action expression) in a given formal language is denoted by *WFF*(respectively, *ACT*).

In an intention structure, a *wff*(or an action expression) may appear repeatedly, and any appearance in different places may result in different effects. In our view, each appearance of a *wff*(an action expression, respectively) should be treated as different intentions. To illustrate this, we consider the following scenario (see Fig.2).

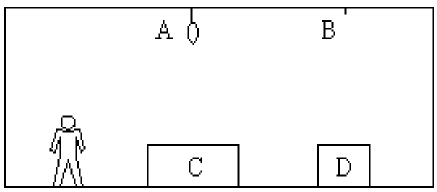

Fig.2 Robot in a blocks world

Example 3. There exists a bulb in the ceiling and two blocks C and D in the room. Robot1 intends to move the bulb from position A to position B. We give two postulates as follows: (1) Robot1 can move only one block at a time; (2) Since the bulb is beyond Robot1's reach, Robot1 has to put the block D on the block C (denoted by *on*(D,C)) in order to fetch the bulb.

It is easy to see that $on(D,C)$ may appear twice (denoted by $on_1(D,C)$ and $on_2(D,C)$, respectively) in his intention structure. We assume that, before Robot1 puts the block D on the block C for reaching down the bulb from A, Robot2 comes and puts the block D on the block C. Suppose Robot1 knows this change, thus, Robot1 should revise his existing intention structure and delete $on_1(D,C)$ from this structure, however, Robot1 should still hold $on_2(D,C)$ as his intentions.

Thus, in order to deal with the change of intentions and give some rationality constraints on the relations between beliefs and intentions, this paper introduce the following concept.

Definition 1. Given WFF and ACT, a function *content*: $IN \rightarrow WFF \cup ACT$ is one if it satisfies the following condition: $|\{j{:}content(j)=i\}|=\aleph_0$, for each $i \in WFF \cup ACT$, where IN is an infinite set called *intention-name* set, $|S|$ denotes the cardinal number of a set S, $\aleph_0$ is the cardinal number of the set of natural numbers.

Based on the above concept, intention structure may be defined formally as follows:

Definition 2. Given IN and *content* as in Definition 1, an intention structure is a tuple $\langle INT, \prec_r, \prec_t \rangle$ consisting of a finite set INT which is a subset of IN, and two binary relations $\prec_r$ and $\prec_t$ defined on the set INT, where $\prec_r$ is an *antisymmetric* relation.

For the sake of convenience, in the following, we regard $j \in INT$ as $content(j)$ unless it is pointed out specially.

Definition 3. Given an intention structure $IS=\langle INT, \prec_r, \prec_t \rangle$, then $\prec_r^* =_{def} t(\prec_r)$, where $t(\prec_r)$ is the transitive closure of $\prec_r$.

Definition 4. We say that an intention structure $\langle INT, \prec_r, \prec_t \rangle$ serves a goal $g \in WFF$ iff the following conditions are satisfied:

(1) $g \in INT$;

(2) $\forall i \in INT(i=g \text{ or } g \prec_r^* i)$.

In this definition, condition (2) says that each of intentions in an intention structure for a goal g is relative to the goal g. From this formal definition and the fact that $\prec_r$ is an *antisymmetric* relation, it is easy to show the following lemma.

Lemma 1. Let $\langle INT, \prec_r, \prec_t \rangle$ be an intention structure for a goal g, then there is no intention such that $i \prec_r^* g$.

Definition 5. Let $IS_g=\langle INT, \prec_r, \prec_t \rangle$ be an intention structure for a goal g, then $\prec_t$ is *temporal-well* iff it satisfies the following conditions:

(1) $\forall i \in INT\ (i \prec_t g)$;

(2) $\forall i_1,i_2,i_3 \in INT\ (i_1 \prec_t i_2 \text{ and } i_2 \prec_t i_3 \Rightarrow i_1 \prec_t i_3)$;

(3) $\forall i_1,i_2 \in INT\ (i_1 \prec_t i_2 \Rightarrow i_2 \not\prec_t i_1)$.

Definition 6. Let $\prec_t$ be *temporal-well*, we call $\prec_t$ *temporal-complete* if it is a linear order, i.e., for each $i,j \in INT, i \prec_t j$ or $j \prec_t i$.

Definition 7. Let $IS=\langle INT, \prec_r, \prec_t \rangle$ be an intention structure, $S \subseteq INT$ and $S \neq \varnothing$, then the set $first(S, \prec_t)$ and $last(S, \prec_t)$ are defined as follows:

(1) $first(S, \prec_t) =_{\text{def}} \{a{:}a \in S \text{ and } \neg\exists b \in S(b \prec_t a)\}$;

(2) $last(S, \prec_t) =_{\text{def}} \{a{:}a \in S \text{ and } \neg\exists b \in S(a \prec_t b)\}$;

If $\prec_t$ is *temporal-complete* then there is only one element in $first(S, \prec_t)$ or $last(S, \prec_t)$. We also denote this element by $first(S, \prec_t)$ (respectively, $last(S, \prec_t)$). In order to abbreviate the formulas, we shall omit $\prec_t$ from $first(S, \prec_t)$ and $last(S, \prec_t)$.

Definition 8. Let $IS=\langle INT, \prec_r, \prec_t \rangle$ be an intention structure for the goal g, an intention *trace* of IS is a sequence of intentions $i_1 i_2 i_3 \ldots i_n$ satisfying the following conditions:

(1) $i_1=g$;

(2) $\forall_{1\le k\le n}(i_k \prec_r i_{k+1})$.

In the following, $trace(IS_g)$ denotes the set of all *traces* of *IS*.

Definition 9. If $l_1,l_2\in trace(IS)$, $l_1=i_1i_2i_3\cdots i_n$ and $l_2=i'_1\, i'_2\, i'_3 \ldots i'_m$ $(n<m)$ such that $i_j=i'_j$ $(1\le j\le n)$, then we say l_2 properly contains l_1 and denoted by $l_1\subset l_2$.

Definition 10. Let $IS=\langle INT,\prec_r,\prec_t\rangle$ be an intention structure and $i\in INT$, then $BAS_{IS}(i)$ is a sequence of intentions $i_1i_2i_3\ldots i_n$ satisfying the following conditions:

(1) $\{i_1,i_2,i_3\ldots i_n\}=\{j{:}j\in INT \text{ and } j\prec_t i\}$;

(2) $\forall_{1\le k\le p\le n}(i_p \not\prec_t i_k)$.

In order to abbreviate the formulas, we shall omit *IS* from BAS_{IS}.

Definition 11. Let $IS=\langle INT,\prec_r,\prec_t\rangle$ be an intention structure and $i\in INT$, then $BAS\text{-Leaf}_{IS}(i)$ is a sequence of intentions $i_1i_2i_3\cdots i_n$ satisfying the following conditions:

(1) $\{i_1,i_2,i_3\ldots i_n\}=\{j{:}j\in INT \text{ and } j\prec_t i \text{ and } j_{\prec_r}=\varnothing\}$;

(2) $\forall_{1\le k\le p\le n}(i_p \not\prec_t i_k)$.

In order to abbreviate the formulas, we shall omit *IS* from BAS_{IS}.

Definition 12. Let $IS=\langle INT,\prec_r,\prec_t\rangle$ be an intention structure and $i\in INT$, then $i_{\prec_r^*}$ and $i_{\prec_r}$ are defined as follows:

(1) $i_{\prec_r^*}=_{\text{def}}\{j{:}j\in INT \text{ and } i\prec_r^* j\}$;

(2) $i_{\prec_r}=_{\text{def}}\{j{:}j\in INT \text{ and } i\prec_r j\}$.

Definition 13. A *normal intention structure* $IS_g=\langle INT,\prec_r,\prec_t\rangle$ for a goal g is an intention structure which satisfies the following conditions:

(1) $\prec_t$ is *temporal-complete*;

(2) $\forall i\in INT\,(i_{\prec_r}\cap ACT\ne\varnothing\Rightarrow|i_{\prec_r}|\le 2)$;

(3) $\forall i\in INT\,(i_{\prec_r}\cap ACT=\{a\} \text{ and } |i_{\prec_r}|=2\Rightarrow i_{\prec_r}\cap wff\ne\varnothing \text{ and } i_{\prec_r}\cap wff=SAC(a))$, where $SAC(a)$ is the preconditions of the action a;

(4) $\forall i\in INT(i\ne g\Rightarrow\exists!j(j\in INT \text{ and } j\prec_r i))$.

As mentioned above, we need to represent agent's predictions of the results caused by the following events: (1) A property becomes true in the world; (2) An action has been executed in the world.

In this paper, we introduced the notation $pred(B,\alpha)$, where B is a belief set and α is a *wff* or action expression. Intuitively, $pred(B,\alpha)$ represents agent's prediction of the likely outcomes after the action α is executed or the property α becomes true in the world characterized by the belief set B. $pred(B,\alpha)$ may be regarded as a belief set concerning future. We neglect that how $pred(B,\alpha)$ is obtained. It is beyond the scope of this paper and belongs to the realm of reasoning about action and change, which has been a focus of research in AI for many years. The framework presented in this paper can be conjoined with whatever theory of action and change, which is appropriate for a given task.

Definition 14. Let $\alpha_1\alpha_2\ldots\alpha_n$ be a sequence of *wffs* and actions, and B be a belief set, then $pred(B,\alpha_1\alpha_2\ldots\alpha_n)$ is defined inductively as follows:

(1) $pred(B,\varnothing)=_{\text{def}}B$;

(2) $pred(B,\alpha_1\alpha_2\ldots\alpha_n)=_{\text{def}}pred(pred(B,\alpha_1\alpha_2\ldots\alpha_{n-1}),\alpha_n)$.

Definition 15. Let $IS=\langle INT,\prec_r,\prec_t\rangle$ be an intention structure for a goal g and $I_0\subseteq INT-\{g\}$, then $IS'=IS\theta I_0$ is an intention structure defined as follows:

(1) $INT'=_{\text{def}}INT-\{i\in INT{:}\exists j\in I_0(j\prec_r^* i) \text{ or } i\in I_0\}$;

(2) $\prec'_t =_{def} \prec_t \Downarrow INT'$, $R\Downarrow A$ denotes the restriction of a relation R w.r.t A;

(3) $\prec'_r =_{def} \prec_r \Downarrow INT'$.

It is easy to show that, if IS is a normal intention structure then $IS\theta I_0$ is also a normal intention structure.

2.3 Axioms on the intention structures

In this subsection, we will present some axioms on the normal intention structures based on the above basic concepts.

Given a normal intention structure $IS=\langle INT, \prec_r, \prec_t\rangle$ and the Belief set Be, if IS satisfies the following axioms 1-8, then we say that IS is *well* with respect to the belief set Be.

Axiom 1. $\exists i\in INT \forall j\in INT(i\in WFF$ and $(i=j$ or $i\prec_r^* j))$.

This axiom says that every intention structure must serve a goal and each of the intentions in IS must originate from the goal. In the sense of mathematics, axiom 1 states that there exists a least element in the strict partial order structure $\langle INT, \prec_r^* \rangle$. In the following, $Top(IS)$ denotes this element and we say that IS is an intention structure for $Top(IS)$.

Axiom 2. $\forall i,j\in INT(i\prec_r j \Rightarrow j\prec_t i)$.

The above axiom says that, if the intention j is one of steps taken to achieve the intention i, then j should be scheduled to be achieved before i.

Axiom 3. $\forall i\in INT \forall j\in INT- i_{\prec_r^*}\ \neg\exists i_1,i_2\in\{i\}\cup i_{\prec_r^*}\ (j\prec_t i_1$ and $i_2\prec_t j)$.

Axiom 3 is not essential to modeling intentions. In order to simplify our treatment, this paper presents this axiom which states that there doesn't exist concurrency in the intention structures, that means, there is no intention which is unrelated to the intention i and is scheduled to be achieved in the course of achieving i.

Axiom 4. $\forall i\in INT(\ i_{\prec_r} \cap ACT \neq\varnothing \wedge |\ i_{\prec_r}\ |=2 \Rightarrow p\prec_t a)$, where $i_{\prec_r}=\{p,a\}, a\in ACT$.

By the definition of normal intention structure and the above axiom, we know that a precondition of an action is always planned to be achieved before executing this action

Axiom 5. $\forall i,j\in INT\cap WFF(content(i)=content(j)\Rightarrow\neg\exists l\in trace(IS)(i\in l$ and $j\in l))$.

Intuitively, this axiom means that there don't exist two intentions i and j which have the same expression and one serves the achievement of another.

Note, since the executing of the same action in different worlds may lead to different effects, if $content(i)$ and $content(j)$ are the same action then it is possible that $\exists l\in trace(IS)(i\in l$ and $j\in l)$.

Definition 16. Let $IS=\langle INT, \prec_r, \prec_t\rangle$ be a normal intention structure and $S\subseteq INT$, then $\Delta(S, \prec_t)$ is a sequence of intentions $i_1 i_2 i_3 \ldots i_n$ satisfying the following conditions:

(1) $\{i_1 i_2 i_3 \ldots i_n\}=S$;

(2) $i_j \prec_t i_{j+1}$, for each $1\leq j<n$.

Axiom 6. $\forall i\in INT-\{Top(IS)\}(\neg\exists S\subseteq i_{\prec_r}\ (i\in pred(B,\Delta(\ i_{\prec_{r,1}}, \prec_{t,1}))$ and $S\neq\varnothing))$, where $B=pred(Be,BAS\text{-}Leaf_{IS1}(first(\ i_{\prec_{r,1}^*})))$; $IS1=\langle INT1, \prec_{t,1}, \prec_{r,1}\rangle=IS\theta S$.

The above axiom sates there is no superfluous intentions in intention structures. In other words, all intentions in the intention structure IS are necessary for the achievement of the goal $Top(IS)$. In particular, it is easy to know that, if an agent has the belief set B and a *wff* $p\in B$, then there doesn't exist a normal intention structure IS for p which is *well* with respect to B. In other words, any goal must not be hold with respect to the current belief set.

In our framework, if an intention is an action, then achieving this intention just means that agent executes this action, thus, the action need not to be divided to finer intentions. So, we present the following axiom.

Axiom 7. $\forall i\in INT(i\in ACT\Rightarrow i_{\prec_r}=\varnothing)$.

Axiom 8. $\forall i\in WFF\cap INT\,(i_{\prec_r}\neq\varnothing\Rightarrow i\in pred(pred(Be,BAS\text{-}Leaf(first(i_{\prec_r^*}))),S))$, where $S=i_1i_2i_3\ldots i_n$ such that:

(1) $\{i_1i_2i_3\ldots i_n\}=\{j: i\prec_r^* j$ and $j_{\prec_r}=\varnothing\}$;

(2) $i_j\prec_t i_{j+1}$, for each $1\le j<n$.

Axiom 8 reveals the logical constraints on the relations between the upper-level intentions and the lower-level intentions. The realization of the lower-level intentions leads to that of the corresponding upper-level ones.

Lemma 2. $IS=\langle INT,\prec_r,\prec_t\rangle$ is a normal intention structure which is *well* with respect to the belief set B, then

(1) $\forall i,j\in INT(i\prec_r^* j\Rightarrow j\prec_t i)$;

(2) $\forall i\in INT\ (BAS(i)=BAS(last(i_{\prec_r^*}))\circ last(i_{\prec_r^*}))$, where $\circ$ denotes with the combination of two intention sequences, more precisely, if $l_1=i_1i_2i_3\cdots i_n$ and $l_2=j_1j_2j_3\ldots j_m$ then $l_1\circ l_2=i_1i_2i_3\ldots i_n j_1j_2j_3\ldots j_m$;

(3) $\forall S\subseteq INT\forall i\in INT(first(S)=first(i_{\prec_r^*})$ and $last(S)=last(i_{\prec_r^*})\Rightarrow S\subseteq i_{\prec_r^*})$;

(4) $\forall l\in trace(IS)\forall i,j\in l(\{content(i),content(j)\}\cap WFF\neq\varnothing\Rightarrow$
$(i\prec_r^* j\Rightarrow\neg\exists i',j'\in l(content(i)=content(i')$ and $content(j)=content(j')$ and $j'\prec_r^* i')$;

(5) $\forall i,j\in INT(i\prec_t j\Leftrightarrow last(i_{\prec_r})\prec_t last(j_{\prec_r}))$;

(6) $\forall a\in INT\exists! i\in WFF\cap INT\,(a\in ACT\Rightarrow i\prec_r a)$;

(7) $\forall i,j\in INT(i\prec_t j\Rightarrow\forall k\in i_{\prec_r^*}(k\prec_t j))$.

Proof. (1) Obviously.

(2) It is easy to know that $BAS(last(i_{\prec_r^*}))\circ last(i_{\prec_r^*})\subseteq BAS(i)$. In the following, we show $BAS(i)\subseteq BAS(last(i_{\prec_r^*}))\circ last(i_{\prec_r^*})$. Suppose that there exists $a\in BAS(i)$ such that $a\notin BAS(last(i_{\prec_r^*}))\circ last(i_{\prec_r^*})$. So, $a\neq i$ and $a\notin i_{\prec_r^*}$, moreover, since the order $\prec_t$ is *temporal-complete*, we get $last(i_{\prec_r^*})\prec_t a$ and $a\prec_t i$, contradicting Axiom 3. Hence, (2) holds.

(3) Let $S\subseteq INT$ and $i\in INT$ such that $first(S)=first(i_{\prec_r^*})$ and $last(S)=last(i_{\prec_r^*})$. Suppose that $a\in S$ and $a\notin i_{\prec_r^*}$. So, $a\neq first(S)$ and $a\neq last(S)$. Hence, we get $a\prec_t last(S)$ and $a\prec_t last(i_{\prec_r^*})$. On the other hand, according to temporal-completeness and $a\in S$, we have $first(S)\prec_t a$(i.e. $first(i_{\prec_r^*})\prec_t a$). A contradiction immediately follows from $a\neq i$ (otherwise, contradicts $last(S)\prec_t i$) and Axiom 3. So $a\in i_{\prec_r^*}$. Hence $S\subseteq i_{\prec_r^*}$.

(4) It immediately follows from Axiom 5.

(5) Let $i,j\in INT$. It is easy to know that, if $i=j$ or $i=last(j_{\prec_r})$, the conclusion holds trivially. In the following, we suppose that $i\neq j$ and $i\neq last(j_{\prec_r})$.

Suppose that $i\prec_t j$ and $last(i_{\prec_r})\not\prec_t last(j_{\prec_r})$. Since $\prec_t$ is *temporal-complete*, we have $last(j_{\prec_r})\prec_t last(i_{\prec_r})$, By Axiom 2, we get $last(i_{\prec_r})\prec_t i$, so $last(j_{\prec_r})\prec_t i$ holds.Since $i\prec_t j$, we get $i\not\prec_r^* j$. By $i\neq last(j_{\prec_r})$ and $last(j_{\prec_r})\prec_t i$, we have $j\not\prec_r^* i$(otherwise, $i\prec_t last(j_{\prec_r})$, a contradiction). Since we have $last(j_{\prec_r})\prec_t i$ and $i\prec_t j$, a contradiction follows from Axiom 3 and $i\not\prec_r^* j$ and $j\not\prec_r^* i$. Hence $last(i_{\prec_r})\prec_t last(j_{\prec_r})$ holds.

Suppose that $last(i_{\prec_r})\prec_t last(j_{\prec_r})$ and $i\not\prec_t j$. Since $\prec_t$ is *temporal-complete*, we have $j\prec_t i$. By the above proof, we get $last(j_{\prec_r})\prec_t last(i_{\prec_r})$, a contradiction. So $i\prec_t j$.

(6) It immediately follows from Definition 13 and Axiom 7.

(7) Suppose that $i\prec_t j$, $k\in i_{\prec_r^*}$ and $k\not\prec_t j$. Since $\prec_t$ is *temporal-complete*, we have $j\prec_t k$. Since $i\prec_t j$, so $i\prec_t k$, on the other hand, by (1), we get $k\prec_t i$, a contradiction. Hence (7) holds. □

Definition 17. A *fine* intention structure IS is a normal intention structure which satisfies the following conditions: $\{i: i\in INT\wedge\neg\exists j(j\in INT\wedge i\prec_r j)\}\subseteq ACT$.

Definition 18. Let IS be a fine intention structure, then the plan $Plan(IS)$ is a sequence of intentions $a_1a_2\ldots a_n$ satisfying the following conditions:

(1) $\{a_1a_2\ldots a_n\}=\{i: i\in INT\wedge\neg\exists j(j\in INT\wedge i\prec_r j)\}$;

(2) $a_i\prec_t a_{i+1}$, $(1\le i\le n-1)$.

Theorem 1(Goal-Intention-Plan theorem). If IS_g is a fine intention structure for the goal g and IS_g is *well* with respect to the belief set B, then $g\in pred(B,Plan(IS_g))$.

Proof. Note that, for any fine intention structure for the goal g, we have $Plan(IS_g)=BAS\text{–}Leaf(g)$, $g\in pred(B, Plan(IS_g))$ immediately follows from Axiom 8. □

3 Discussion

We provide an alternative framework for investigating intentions. The axioms presented in this paper only depict a part of the characteristics of intentions. Other characteristics, such as the compatibility and evolutionary tendency, may be considered further when dealing with the generation and change of intentions. In our views, the researches on intentions may be divided into three categories, i.e. the generation, the change, and the static representation. We have done some work in those categories based on the framework presented in this paper. The work concerning the change of intention structures will appear in another paper.

References:

[1] Wooldridge MJ. Intelligent agents: The key concepts. In: Marik V, Stepankova O, Krautwurmova H, eds. Multi-Agent-Systems and Applications. Berlin: Springer-Verlag, 2002. 3~43.

[2] Georgeff MP, Pell B, Pollack M, Tambe M, Wooldrideg M. The belief-desire-intention model of agency. In: Muller JP, Singh M, Rao A, eds. Intelligent Agents V: Theories, Architectures, and Languages. Berlin: Springer-Verlag, 1999. 1~10.

[3] Cohen PR, Levesque HJ. Intention is choice with commitment. Artificial Intelligence, 1990,42(3):213~261.

[4] Rao AS, Georgeff MP. Modeling rational agents within a BDI-Architecture. In: Fikes R, Sandewall E, eds. Proc. of the Knowledge Representation and Reasoning. Cambridge: Morgan Kaufmann Publishers, 1991. 473~484. http://citeseer.nj.nec.com/122564. html

[5] Konolige K, Pollack ME. A representationalist theory of intention. In: Bajcsy R, ed. Proc. of the 13th Int'l Joint Conf. on Artificial Intelligence. Cambridge: Morgan Kaufmann Publishers, 1993. 390~395.

[6] Caredon L, Padgham L, Rao A, Sonenberg E. Revisiting rationality for agents with intentions. In: Yao X, ed. Proc. of the 8th Australian Joint Conf. on Artificial Intelligence. Singapore: World Scientific Publishers, 1995. 131~138.

[7] Wobcke W. Modeling PRS-Like Agent's mental states. In: Ishizuka M, Sattar A, eds. Trends in Artificial Intelligence. Berlin: Springer-Verlag, 2002. 138~147.

[8] Schild K. On the relationship between BDI logics and standard logics of concurrency. Autonomous and Multi-Agent Systems, 2000,3(3):259~283.