

# 避免逻辑全知的 BDI 语义\*

程显毅<sup>1</sup>, 石纯一<sup>2</sup>

<sup>1</sup>(江苏大学 计算机科学与通信工程学院,江苏 镇江 212013);

<sup>2</sup>(清华大学 计算机科学与技术系,北京 100084)

E-mail: chengxianyi@263.net

http://www.jsust.edu.cn

**摘要:** BDI(belief, desire, intention)是基于 Agent 计算的理论模型,BDI 语义关系着 Agent 计算的发展.通过把相信划分为主观相信和客观相信,把可能世界理解为认知的不同阶段,给出具有进化特征的 BDI 语义.该语义既能描述 Agent,又避免了“逻辑全知”问题.

**关键词:** Agent 计算;BDI;语义;模态算子;认知

**中图法分类号:** TP181 **文献标识码:** A

Agent 是近几年来计算机科学领域中的一个重要概念,这一概念已被广泛应用于 DAI、软件工程、网络等领域.虽然关于 Agent 目前仍没有一个统一的定义,但一个 Agent 可以被看做是拥有信念(belief)、愿望(desire)、意图(intention)等精神状态的意识系统(简称 BDI 模型).采用 Kripke 可能世界语义模型解释 BDI,会存在致命的“逻辑全知”问题.根源在于模态算子 B 的非正规特性.正规模态算子和非正规模态算子的本质区别取决于自然化规则 N 是否有效<sup>[1]</sup>.因为 Agent 相信所有有效公式,这构成了著名的“逻辑全知”问题.为了避免“逻辑全知”问题的出现,许多学者已给出了 BDI 的语义,但大都存在一定的局限性.本文通过把相信  $\phi$  划分为主观相信(仍记为  $B\phi$ )和客观相信(记为  $K\phi$ ,此时称  $\phi$  为知识),把可能世界理解为认知的不同阶段,把 BDI 解释为在某一认知阶段的认知,新的 BDI 语义既能描述 Agent,又避免了“逻辑全知”问题.

## 1 现有 BDI 语义的局限性

现有 BDI 语义有如下几种趋势:

(1) 标准可能世界方法.它用可能世界语义模型新的可达关系解释非正规模态算子<sup>[1]</sup>.

$$V(\alpha, w) \begin{cases} 1, & w \in W, \text{且对 } \forall w', \text{若 } R ww', \text{则 } V(\alpha, w')=1 \\ 0, & \text{否则} \end{cases}$$

符号  $\alpha$  表示非正规模态算子;  $\beta$  表示正规模态算子;  $W$  表示可能世界集合;  $R$  或  $R'$  分别表示非正规模态算子和正规模态算子的可达关系(下同).

标准可能世界语义对认知的解释是,要么一开始就认知了,要么就永远不能被认知,所以标准可能世界与重言式世界等价.也就是说,在标准可能世界语义下,  $\alpha$  和  $\beta$  没有本质区别,所以不能避免“逻辑全知”.

(2) 信念演绎模型.其指导思想是通过公理方法避免在系统中出现逻辑闭包. Levesque 的信念逻辑系统从语义上阻断了逻辑闭包的生成途径, Konolige 的信念演绎模型则是从语法上来阻断逻辑闭包的生成,但二者都

\* 收稿日期: 2000-06-19; 修改日期: 2001-05-09

基金项目: 国家自然科学基金资助项目(69773026)

作者简介: 程显毅(1956 - ),男,黑龙江哈尔滨人,副教授,主要研究领域为多 Agent 系统;石纯一(1935 - ),男,河北秦皇岛人,教授,博士生导师,主要研究领域为人工智能应用基础.

只是给出了有关信念逻辑,而没有考虑有关知识的逻辑,这对于一个能体现人类认知的 Agent 来说是不够完善的.文献[2]给出了认知逻辑 EL,只是在演绎模型中扩充了“知识”模态算子,能很好地解释 B,但未能解释 D 和 I.

(3) 非正规可能世界方法.它承认在某一世界中,存在不能被 Agent 认知的客体.把不能判定真假的可能世界集合记为  $A$ ,称为非正规可能世界.所以有

$$V(\alpha, w) \begin{cases} 1, & w \in W, \text{且对 } \forall w', \text{若 } R ww', \text{则 } V(\alpha, w')=1 \\ 0, & w \in W, \text{且对 } \forall w', \text{若 } R ww', \text{则 } V(\alpha, w')=0 \\ x, & w \in A, x=1 \text{ 或 } x=0 \end{cases}$$

该方法虽然能避免“逻辑全知”问题,但需将可能世界划分为可以认知的客体世界和不能判断真假的客体世界,这种划分在实际系统是无法实现的.

(4) 非标准可能世界方法.非标准可能世界语义不划分可能世界,而是用两个可达关系确定一个认知,即

$$V(\alpha, w) \begin{cases} 1, & w \in W, \text{且对 } \forall w', \text{若 } R^f ww', \text{则 } V(\alpha, w')=1 \text{ 且对 } \forall w'', \text{若 } R^f ww'', \text{则 } V(\alpha, w'')=0 \\ 0, & \text{否则} \end{cases}$$

$R^f, R^f$  是  $W \times W$  子集,分别为重言可达关系和非正规模态可达关系.该语义不增加非正规可能世界,当  $R^f = \emptyset$  时,该语义退化为正规模态系统的 Kripke 解释,使得正规逻辑和非正规逻辑得到了统一.在该语义解释下,自然化规则  $N$  不成立,从而不存在“逻辑全知”问题<sup>[3]</sup>.但两种可达关系的引用有些不自然.

现有的 BDI 语义解释无论是哪一种,都不具有进化特征,这与人的认知过程是相违背的.下一节我们将讨论具有认知进化特征的 BDI 语义.

## 2 具有进化特征的 BDI 语义

### 2.1 定义

现有的 BDI 语义是对进化的省略,对于简单的计算来说,具有简捷和抽象度高的优点,但是,这并不适合处理复杂计算,究其原因,有以下几点: 对于在计算中占大多数的开放性问题来说,进化特征是语义刻画的重要内容,这里并不一定存在一个最终的执行状态,而且计算结果依赖于交互的时刻. 即使对于封闭计算,渐进认知特性也是不容忽视的语义描述内容,现有的 BDI 语义框架由于简化了认知的进程,不但造成理论和实际脱节,也大大减弱了 BDI 语义表达能力,从而缩小了应用范围. 省略进化特征难以反映 Agent 的社会性,不符合 Agent 引入的初衷.

从形式上看,描述 Agent 的逻辑系统就是在古典逻辑的基础上增加主观相信算子  $K$  和客观相信算子  $B$ . 设  $L$  为通常模态逻辑系统所用的语言.

定义 1.  $F = \langle W, R_b, R_k \rangle$  是一个框架,当且仅当  $W$  是非空升序集合,  $R_b, R_k$  分别是  $W$  上知识关系和信念关系(统称为认知关系).

$W$  中的每一个元素可理解为认知过程的一个认知阶段,对于  $w_1 < w_2 \in W$ ,表明认知阶段  $w_2$  高于认知阶段  $w_1$ ,二元关系  $(\varphi, w) \in R_b, (\varphi, w) \in R_k$  分别反映了知识和信念在发展演化过程中各个认知阶段之间的关系.

定义 2. 设  $F = \langle W, R_b, R_k \rangle$  是任一框架,  $V$  是框架  $F$  上对  $L$ -公式( $Form(L)$ )的一个真值赋值( $V: Form(L) \times W \rightarrow \{0, 1\}$ ),当且仅当对  $\forall \alpha, \beta \in Form(L)$  和  $w \in W$  满足如下赋值规则:

$$\begin{array}{l} [V\neg] \\ [V\alpha \rightarrow \beta] \\ [VB] \end{array} \quad V(\alpha, w) \begin{cases} 1, & V(\alpha, w)=0 \\ 0, & \text{否则} \\ 1, & V(\alpha, w)=0 \text{ 或 } V(\beta, w)=1 \\ 0, & \text{否则} \\ 1, & \alpha_b(w) \in R_b, \text{其中 } \alpha_b(w) = (\alpha, w) \\ 0, & \text{否则} \end{cases};$$

$$[VK] \quad V(K\alpha, w) \begin{cases} 1, & \text{当 } \alpha_b(w) \in R_b, \alpha_k(w) \in R_k \text{ 有 } \alpha_k(w) = \alpha_b(w) \\ 0, & \text{否则} \end{cases}$$

$\alpha_k(w)$ 和 $\alpha_b(w)$ 分别表示在认知阶段  $w$  的知识和信念,在阶段  $w$  能断定所相信的信念  $\alpha$  不是知识,或能够断定信念  $\alpha$  不能被同化为知识,则  $B\alpha$  取 0.信念可以同化为知识是指,从公理出发可以演绎出该信念.

根据以上规则和  $\cap, \cup, \leftrightarrow, D, I$  的语法定义,不难得到相应的赋值规则  $[V\cap], [V\cup], [V\leftrightarrow], [VD], [VI]$ .

定义 3.  $M = \langle W, R_b, R_k, V \rangle$  是模型,当且仅当  $F = \langle W, R_b, R_k \rangle$  是一个框架,  $V$  是  $F$  上的一个真值赋值.

定义 4. 设  $M = \langle W, R_b, R_k, V \rangle$  是一个模型,  $\alpha \in \text{Form}(L)$ .

- (1) 对  $w \in W$ , 若  $V(\alpha, w) = 1$ , 记为  $M \models_w \alpha$ , 称  $\alpha$  在  $w$  上有效. 否则称  $\alpha$  在  $w$  上无效.
- (2) 对  $\forall w \in W$ , 若  $V(\alpha, w) = 1$ , 都有  $M \models_w \alpha$ , 称  $\alpha$  在  $M$  上有效, 记为  $M \models \alpha$ .
- (3) 对  $\exists w \in W$ , 若  $V(\alpha, w) = 1$ , 有  $M \models_w \alpha$ , 称  $\alpha$  在  $M$  上可满足. 否则, 称  $\alpha$  在  $M$  上不可满足.

## 2.2 应用

### 2.2.1 BDI 之间的关系

在新的 BDI 语义下有如下有效公式, 它们的证明需要建立 Agent 计算模型, 限于篇幅, 证明从略, 只给出公式的直观意义.

- (1)  $\varphi \rightarrow B\varphi$  重言式本身应该被相信.
- (2)  $B\varphi \rightarrow BB\varphi$  信念本身应该被相信.
- (3)  $\neg B_{\text{false}}$  一个理性认知主体不能相信一个实际上为假的命题.
- (4)  $K\varphi \rightarrow B\varphi$  知识本身是被相信的.
- (5)  $K\varphi \rightarrow \varphi$  只有真的事实才能被知道, 表明  $K$  与真假有联系.
- (6)  $K\varphi \rightarrow KK\varphi$  一个理性认知主体对它知道什么是完全清楚的.
- (7)  $\neg K\varphi \rightarrow K\neg K\varphi$  一个理性认知主体对它不知道什么是完全清楚的.
- (8)  $\neg K_{\text{false}}$  一个理性认知主体不知道一个实际上为假的命题.
- (9)  $KK\varphi \rightarrow K\varphi$  知道的知识也是知识.
- (10)  $K\varphi \rightarrow BK\varphi$  知道本身是被相信的.
- (11)  $(K\varphi \vee \neg K\varphi) \rightarrow (B\varphi \rightarrow K\varphi)$  当信息完全时, 信念就与知识同化.
- (12)  $K\varphi \rightarrow I\varphi$  曾经做过的实际一定能做.
- (13)  $D\varphi \rightarrow B\varphi$  相信想要做的.
- (14)  $I\varphi \rightarrow B\varphi$  相信实际能做的.
- (15)  $I\varphi \wedge B(\neg\varphi)$  不允许意图、信念的不相容性.
- (16)  $I\varphi \wedge (\neg B\varphi)$  不允许意图的不完全性.

### 2.2.2 主观相信和客观相信

“逻辑全知”问题会产生一些致命的逻辑结论和结果.

首先, 它要求 Agent 的信念集不能含有矛盾的命题和公式, 而人的信念在逻辑上却不需要(也不可能)保持一致, 也不在意这种不一致信念的存在. 原因在于人们对信念的相信区别为主观相信(B)和客观相信(K). 早期研究信念是不区分 B 和 K 的, 但二者有明显的区别, 知识可以看做是经过验证为真的信念, 信念是可以撤消、修正的知识; 知识是指在主观和客观上都能充分地承认其为真的判断, 而信念与真假却没有联系, 因为我们完全可能会相信一个实际上为假的命题. 所以信念提供的不仅是知识, 也包含谬误, 人类的认知过程就是由信念和借助推理行为从一个信念到另一个信念的构筑过程; 知识具有传递性, 即可以从一个认知主体传递到另一个主体, 而信念没有这种传递性<sup>[4]</sup>.

其次, 知识和信念在重言式和逻辑结果意义上封闭还可以产生逻辑全知的其他途径, 如在合取、重言蕴涵、逻辑等价意义上的封闭<sup>[5]</sup>. 如  $\varphi \vee \neg\varphi$  有效, 不能得到  $K\varphi \vee \neg K\varphi$  有效,  $K\varphi \vee \neg K\varphi$  不有效, 表明 Agent 的信息是不完全的, 当信息完全时:

$$(K\varphi \vee \neg K\varphi) \rightarrow (B\varphi \rightarrow K\varphi).$$

此时信念转化为知识,或知识和信念同化.另外,组合原则遭到破坏,组合原则用在真值方面是说,公式的真值由其子公式以及相应的复合方式组成.

在 Agent 计算中,  $B$  表示相信可以做的,是 Agent 计算的前提.区分了  $B$  和  $K$ ,并把  $w \in W$  作为认知过程的一个阶段,我们有如下  $B$  的语义:

$$V(B\alpha, w) \begin{cases} 1, & w \in W, \text{ 且对 } \forall w' \in W, w < w', V(K\alpha, w') = 1 \text{ (信念 } \alpha \text{ 被同化为知识)} \\ 0, & \text{ 否则} \end{cases}$$

信念  $B$  的语义是说,在阶段  $w$  能断定所相信的信念  $\alpha$  不是知识,或能够断定信念  $\alpha$  不能被同化为知识,则  $B\alpha$  取 0,否则  $B\alpha$  取 1.

### 2.2.3 意图 $I$ 的语义

$I$  将影响和约束 Agent 的未来行为,它是 Agent 计算的起因,即 Agent 将根据意图作出行为决策、选择规划、执行动作. $I$  的基本内涵是对未来行为的理性选择,现有的意图  $I$  的语义大多基于可能世界间的可达关系,即将  $I$  视为 Agent 对可能世界的选择,其不足是存在副作用,如  $I\alpha$  表示有意图拔牙,  $I\beta$  表示有意图牙痛,  $I(\alpha \rightarrow \beta)$  表示拔牙必牙痛.虽然一个拔牙患者知道  $I\alpha, I(\alpha \rightarrow \beta)$  为真,但他不在意  $I\beta$  为假,而使用可能世界语义模型恰恰得到  $I\beta$  为真.也就是说,虽然有  $I\beta \rightarrow I(\alpha \rightarrow \beta)$ ,但没有  $I\alpha \rightarrow I\beta$ .我们将  $I$  视为对信念的修正:

$$V(I\alpha, w) \begin{cases} 1, & w \in W, \text{ 且对 } \forall w' \in W, w' < w, V(B\alpha, w') = 1 \text{ (对信念 } \alpha \text{ 已修正过,或不需要修正)} \\ 0, & \text{ 否则 (对信念 } \alpha \text{ 无法修正,或没有完成修正)} \end{cases}$$

意图  $I$  的语义是说,如果需要修正的信念都能修正,则持续该意图.信念  $\alpha$  无法修正或没有完成修正,有可能是资源约束造成的.

### 2.2.4 愿望 $D$ 的语义

在 Agent 计算中,  $D$  表示想要做的,是 Agent 计算的目标,我们将  $D$  视为  $D$  和  $B$  之间距离的度量.

$$V(D\alpha, w) \begin{cases} 1, & w \in W, \text{ 且对 } \forall w' \in W, w < w', V(I\alpha, w') = V(I\alpha, w) \\ 0, & \text{ 否则} \end{cases}$$

愿望  $D$  的语义是说,如果与更高认知阶段有相同的意图,则继续执行规划,否则修正信念.

由于我们采用线性时序和信念修正机制,随着认知阶段的逐步深入,认知或得到加强、或得到修正、或得到删除.

## 2.3 在新的语义下可以避免“逻辑全知”

正规模式逻辑系统应有  $K$  公理和  $N$  规则,对一般非正规模式逻辑系统  $K$  公理也是该系统的定理,只是对某些算子  $N$  规则可能不成立.下面我们证明在新的语义下,  $K$  公理成立,而  $N$  规则不成立.

**定理 1.** 规则  $N(\varphi \rightarrow K\varphi)$  对模型  $M = \langle W, R_b, R_k, V \rangle$  不有效.

证明:因为  $\varphi$ ,所以存在  $w' \in W$  使  $V(\varphi, w') = 1$ .从人类的认知经验可知,为真的事实,在某一认知阶段没有得到证实的情况下,不会是客观相信,至多是主观相信,所以  $\varphi_k(w) = \varphi_b(w)$  不成立,  $K\varphi = 0$ ,从而  $N$  规则不成立.

**定理 2.** 认知从低级阶段  $w_1$  跃入高级阶段  $w_2$ ,当且仅当在  $w_1$  阶段主观相信与客观相信是一致的.  $w_1 \in W, w_2 \in W, w_1 < w_2, \alpha \in \text{Form}(L)$ ,若  $\alpha_b(w_1) \in R_b$ ,则  $\alpha_k(w_2) \in R_k$ .

**定理 3.** 高级阶段的客观相信都是低级阶段的主观相信.  $w_1 \in W, w_2 \in W, w_1 < w_2, \alpha \in \text{Form}(L)$ ,若  $\alpha_k(w_2) \in R_k$ ,则  $\alpha_b(w_1) \in R_b$ .

定理 2 和定理 3 的直观意义很清楚,证明从略.

**定理 4.** 公理  $K(B(\alpha \rightarrow \beta) \rightarrow (B\alpha \rightarrow B\beta))$  对模型  $M = \langle W, R_b, R_k, V \rangle$  有效.

证明:  $K$  公理形式上相当于人们认知事物所使用的三段论.

因为  $B(\alpha \rightarrow \beta) = 1, B\alpha = 1$ ,

所以存在  $w_1 \in W, w_2 \in W$  使得  $(\alpha \rightarrow \beta, w_1) \in R_k, (\alpha, w_2) \in R_k$ ,取  $w^* = \min(w_1, w_2)$ ,由定理 3 得,  $(\beta, w_1) \in R_k, (\beta, w_1) \in R_b$ ,从

而  $B\beta=1$ .

### 3 结束语

在 Agent 计算中, B 表示相信可以做的, 是 Agent 计算的前提; K 表示曾经做过的, 是 Agent 计算的基础; D 表示想要做的, 是 Agent 计算的目标; I 表示实际能做的, 是 Agent 计算的动机. K, B, D, I 之间在认知过程中的关系形如  $B \xrightarrow[I]{I} D$ .

非正规模态算子 B, D, I 和正规模态算子 P 之间在的关系如图 1 所示.

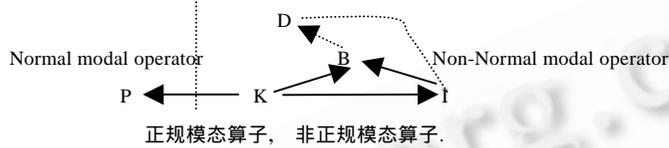


Fig.1 Relations of B, D, I, K, P

图 1 B, D, I, K, P 之间存在的关系

图中箭头表示实质蕴涵, 可以看出 K 是连接非正规模态算子和正规模态算子的桥梁, 为了完成 B 到 D 的计算, 我们有两条路可以走, 第 1 条路是计算 I 到 D, 这是传统的计算模式(实际能做的都是想做的, 计算是封闭的, 不能得到新知); 第 2 条路是借助于 I 和 K, 计算 B 到 D, 使 Agent 信息由不完全向完全进化, 达到相信可以做的, 一定是想要做的, 这是 Agent 计算所希望的模式. 由于篇幅所限, 有许多问题, 如信念的修正、信念与知识的同化以及 Agent 计算模型等将在后续文章中讨论.

#### References:

- [1] Zhou, Bei-hai. Introduction to Modal Logic. Beijing: Beijing University Press, 1997 (in Chinese).
- [2] Yang, Kun, Chen, Jian-zhong, Sun, De-gang, *et al.* Logical omniscience problem in epistemic logic and the methods to resolve it. ACTA Scientiarum Naturalium Universitatis JILINENSIS 3, 1999, 23(3):40~43 (in Chinese).
- [3] Hu, Shan-li, Shi, Chun-yi. A semantic interpretation for agent's non-normal modal operators. Journal of Computer Research and Development, 1999, 36(10):1153~1157 (in Chinese).
- [4] Liu Rui-sheng, Sun Ji-gui, Liu Xu-hua. Epistemic logic (1): a logical framework for knowledge and belief. Chinese Journal of Computers, 1998, 21(7):627~637 (in Chinese).
- [5] Jennings, N.R. Agent-Based computing: promise and perils. In: Proceedings of the International Joint Conference on Artificial Intelligence. Sweden: Stockholm, 1999. 1429~1436. <http://ijcai.org>.

#### 附中文参考文献:

- [1] 周北海. 模态逻辑导论. 北京: 北京大学出版社, 1997.
- [2] 杨鲲, 陈建中, 孙德刚, 等. 认知逻辑中逻辑全知问题及其解决方法. 吉林大学自然科学学报, 1999, 23(3):40~43.
- [3] 胡山立, 石纯一. 适用于 Agent 非正规逻辑模态算子的一种语义解释. 计算机研究与发展, 1999, 36(10):1153~1157.
- [4] 刘瑞胜, 孙吉贵, 刘叙华. 认识逻辑(1): 关于知识和信念的逻辑框架. 计算机学报, 1998, 21(7):627~637.

## BDI Semantic of Avoiding Logical Omniscience Problem\*

CHENG Xian-yi<sup>1</sup>, SHI Chun-yi<sup>2</sup>

<sup>1</sup>(College of Computer Science and Communications Engineering, Jiangsu University, Zhenjiang 212013, China);

<sup>2</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

E-mail: chengxianyi@263.net

<http://www.jsust.edu.cn>

**Abstract:** BDI (belief, desire, intention) is a theoretical model based on agent computing, while its semantic is the key to the development of agent computing. The new BDI semantic is presented to have an evolution character by dividing B into subjective belief and objective belief, and by regarding possible world as a temporal stage of recognition procedure, which can describe agent and avoid the logical omniscience problem.

**Key words:** agent computing; BDI; semantic; modal operator; recognition

\* Received June 19, 2000; accepted May 9, 2001

Supported by the National Natural Science Foundation of China under Grant No.69773026

## 第 9 届 Rough 集、模糊集、数据挖掘与粒度计算国际学术会议(RSFDGrC 2003)

### 征文通知

RSFDGrC 2003 是 RSFDGrC 系列国际会议的第 9 届会议,该系列会议每两年举行一次.RSFDGrC 2003 是该系列会议第一次在中国举办.RSFDGrC 2003 将由国家自然科学基金会、重庆邮电学院等有关单位共同主办,于 2003 年 5 月 26 日~29 日在重庆召开.会议将针对 Rough 集、模糊集、数据挖掘、粒度计算等计算智能的理论与应用进行学术研究与讨论.

#### 一、 征文范围

Rough 集理论及应用	计算智能	机器学习	文字计算
Fuzzy 集理论及应用	粒度计算	软计算及其应用	进化计算
Petri 网	软计算的逻辑基础	非经典逻辑	神经网络
软计算复杂性	空间推理	统计推理	智能 Agent
多标准决策分析	决策支持系统	知识发现与数据挖掘	多 Agent 技术
网络智能	集成智能系统	近似推理与不确定推理	数据仓库
模式识别与图像处理	其他有关领域		

#### 二、 征文要求

- (1) 论文未被其他会议、期刊录用或发表;
- (2) 大会工作语言为英语;
- (3) 来稿一式 4 份;
- (4) 为联系方便,请务必提供作者的姓名、单位、通信地址、电话、传真及 E-mail 地址;
- (5) 论文版面格式及投稿信息请参看会议主页.

#### 三、 重要日期

特邀小组会议申请: 2002 年 11 月 10 日(收到日期)

征文截止日期: 2002 年 12 月 10 日(收到日期)

录用通知日期: 2003 年 1 月 10 日(发出日期)

提交正式论文截止日期: 2003 年 2 月 10 日(收到日期)

#### 四、 联系方式

联系地址: 重庆市南岸区黄桷埡镇堡上园 1 号 重庆邮电学院计算机科学与技术研究所

邮政编码: 400065

联系人: RSFDGrC' 程序委员会王国胤教授

联系电话: 023-62460066

传真: 023-62461882

E-mail: wanggy@cqupt.edu.cn

主页: <http://www.cqupt.edu.cn/rsfdgrc>