

基于 Nearest-Biclusters 协作过滤技术的 效用图结构学习算法

王黎明¹⁾ 李 琨^{1),2)}

¹⁾(郑州大学信息工程学院 郑州 450001)

²⁾(河南职工医学院 郑州 451191)

摘 要 在多议题协商研究中,议题之间的依赖关系增加了协商 Agent 效用函数的复杂性,从而使得多议题协商变得更加困难.基于效用图的多议题依赖协商模型是体现议题间依赖关系的多议题协商模型.在该协商模型中,协商双方仅需要较少的协商步数就能够找到满足 Pareto 效率的协商结局.如何有效地学习买方 Agent 的效用图结构是该协商模型的关键.文中基于 Nearest-Biclusters 协作过滤技术的思想提出了一种新的效用图结构学习算法(NBCFL 算法).该算法首先利用 Nearest-Biclusters 协作过滤技术发现买方偏好的局部匹配特性,提取与当前买方 Agent 类型相同的买方 Agent 所产生的协商历史记录,然后通过计算各议题间的依赖度学习买方 Agent 的效用图结构.实验表明在参与协商的买方 Agent 类型不同的条件下,NBCFL 算法比 IBCFL 算法能更好地学习买方 Agent 的效用图结构.

关键词 效用图;多议题协商;协作过滤;双向聚类

中图法分类号 TP18 **DOI 号:** 10.3724/SP.J.1016.2010.02291

An Algorithms for Learning the Structure of Utility Graph Based on Nearest-Biclusters Collaborative Filtering

WANG Li-Ming¹⁾ LI Kun^{1),2)}

¹⁾(School of Information Engineering, Zhengzhou University, Zhengzhou 450001)

²⁾(Henan Medical College for Staff and Workers, Zhengzhou 451191)

Abstract In the research of multi-issue negotiation, the interdependencies between issues greatly complicates the negotiation agents' utility functions, so this makes negotiation more difficultly. The multi-issue negotiation model based on Utility Graph is the multi-issue dependence negotiation model which considers interdependencies between issues. The negotiants need a few number of negotiation steps to reach Pareto-efficient agreements in the negotiation model. The key problem of the negotiation model is how to learn the structure of Utility Graph effectively. This paper proposes a new algorithm for learning the structure of Utility Graph based on Nearest-Biclusters Collaborative Filtering(NBCFL). Firstly, the algorithm takes advantage of the trait that Nearest-Biclusters Collaborative Filtering could detect partial matching of buyers' preferences, and collects the negotiation data which were produced by buyer that is the same class with active buyer. Secondly, it retrieves the structure of Utility Graph using the degree of interdependencies between issues. The experiments show that NBCFL algorithm can learn the structure of Utility Graph more effectively than IBCFL algorithm on condition that buyers from different classes of buyers.

Keywords utility graph; multi-issue negotiation; collaborative filtering; biclustering

1 引言

多议题协商是多 Agent 系统研究的热点之一。目前关于多议题协商问题有大量的研究^[1-5],其中大部分研究工作都假定协商议题之间相互独立^[1-3]。文献[3-5]是我们的前期工作,在此基础上,本文进一步考虑议题之间的依赖性,这种依赖性增加了多议题协商问题的复杂性。Klein 等人^[6]首次将多议题依赖引入到多议题协商模型中,协商 Agent 使用模拟退火和爬山搜索方法搜索 Pareto 最优解;Takayuki 等人^[7-8]利用带有约束的效用函数表示协商 Agent 的非线性偏好(即议题间存在依赖关系),提出了基于拍卖的多议题依赖协商模型;Lai 等人^[9]在协商 Agent 具有非线性偏好的条件下,利用基于协商基线的协商协议对多议题协商模型进行了研究;Somefun 等人^[10]利用超立方体结构类型的效用函数表示买方 Agent 的非线性偏好,并以绑定(bundling)销售模型为基础对多议题依赖协商问题进行了研究;Robu 等人^[11-12]提出了一种基于效用图的多议题依赖协商模型,该模型采用效用图表示买方 Agent 非线性偏好,并利用图形结构特点大大降低了多议题依赖协商问题的复杂性。文献[12]中的实验表明,随着效用图结构学习准确率的增加,完成协商所需要的协商步随之减少,而协商结局满足 Pareto 最优的比率随之增加。因此,能否正确学习效用图的结构是该协商模型中需要解决的一个关键问题。

协作过滤技术^[12]的主要思想是利用协商历史数据发现买方的偏好并为买方提供购买建议。然而这种建议机制是静态的,买方对该建议仅能够接受、拒绝或者坚持己见。在协商框架内,该建议机制可以被协商过程所取代,增强了灵活性。Robu 等人^[12]基于项目协作过滤技术的思想,提出了效用图结构学习算法(IBCFL 算法)。当买方 Agent 类型相同时,该算法能够得到较好的学习效果。因为 IBCFL 算法是利用整个协商历史数据求解议题之间依赖度的,所以当买方 Agent 类型不同时,该算法不能有效地学习效用图结构。

在买方 Agent 类型不同的条件下,本文对基于效用图的多议题依赖协商模型进行了研究,并借鉴 Nearest-Biclusters 协作过滤技术^[13]的思想提出了效用图结构学习算法(NBCFL 算法),且通过实验验证了算法的可行性与有效性。

2 相关定义

定义 1(协商 Agent 集合). $\Sigma = \Sigma_B + \Sigma_S$ 是协商 Agent 的集合,其中 $\Sigma_B = \{B_1, \dots, B_p\}$ 为买方 Agent 集合, p 为买方 Agent 的个数; $\Sigma_S = \{S_1, \dots, S_q\}$ 为卖方 Agent 集合, q 为卖方 Agent 的个数。

定义 2(议题集合及其值向量). $I = \{I_1, I_2, \dots, I_n\}$ 表示协商议题集合,其中 n 表示协商议题的个数, $\mathbf{x} = (x_1, \dots, x_n)$ 表示关于议题集合 I 的一个值向量,其中 $\forall I_i \in I (1 \leq i \leq n)$, $x_i \in X_i$ 表示议题 I_i 的一个取值, X_i 表示议题 I_i 取值的值域。

定义 3(一般联加独立效用函数)^[14]. 设 $Dom(I) = \prod_{i=1}^n X_i$ 表示值向量的状态空间,议题簇集合 $C = \{C_1, \dots, C_r\}$ 是集合 I 的 r 个子集组成的集合且 $\bigcup_{i=1}^r C_i = I$, $\forall i, 1 \leq i \leq r$, 设 $Dom(C_i) = \prod_{j \in C_i} X_j$. 效用 $u(\cdot)$ 是关于 $Dom(C_i)$ 的一般联加独立效用函数,当且仅当存在子效用 $u_i: Dom(C_i) \rightarrow \mathbb{R}$ 使得 $\forall \mathbf{x} \in Dom(I)$, $u(\mathbf{x}) = \sum_{i=1}^r u_i(\mathbf{x}_{C_i})$, 其中 $\mathbf{x}_{C_i} \in Dom(C_i)$ 表示议题簇 C_i 关于值向量 \mathbf{x} 构成的子值向量。

定义 3 中的每一个议题簇 C_i 中包含的议题表示这些议题间存在依赖关系,而子效用 u_i 则反映了这些议题间的依赖度。当 $|C_i| = 1 (1 \leq i \leq r)$ 时,子效用值 u_i 表明购买一个商品时对该商品的评估值;当 $|C_i| > 1$ 时,子效用值 u_i 表明购买两个或两个以上商品的联合效用值,如果联合效用值为正,则表示这些商品之间存在互补性,否则存在可替代性。

定义 4(效用图). 设 $Dom(I) = \prod_{i=1}^n X_i$, 议题簇集合 $C = \{C_1, \dots, C_r\}$ 是集合 I 的 r 个子集组成的集合且 $\bigcup_{i=1}^r C_i = I$. 设 $\forall \mathbf{x} \in Dom(I)$, 则表示效用 $u(\mathbf{x}) = \sum_{i=1}^r u_i(\mathbf{x}_{C_i})$ 的效用图是一个无向图 $G = (I, F)$, 当且仅当满足

- (1) $I = \{I_1, \dots, I_n\}$ 是无向图 G 的点集;
- (2) $\forall (I_i, I_j) \in F, \exists C_s \subset C (1 \leq s \leq r)$, 使得 $I_i, I_j \in C_s$.
- (3) $\forall C_s \subset C (1 \leq s \leq r)$, 效用 $u(\cdot)$ 中的各子效用 u_s 与之关联。

例 1. 设 $I = \{I_1, I_2, I_3, I_4\}$, 议题簇集合 $C = \{\{I_1\}, \{I_1, I_2\}, \{I_2, I_3\}, \{I_2, I_4\}, \{I_3, I_4\}\}$. 设 $C_2,$

C_4 对应的子效用 u_2, u_4 取值(即议题间依赖度)为正, C_3, C_5 对应的子效用 u_3, u_5 取值为负, 则效用图如图 1 所示.

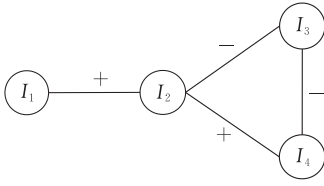


图 1 效用图(+表示互补性, -表示可替代性)

定义 5(买方 Agent 类型集合). $C_A = \{C^1, \dots, C^t\}$ 表示买方 Agent 类型的集合, 其中 t 为买方 Agent 类型的个数; $C^i = \{C^i_1, \dots, C^i_r\}$ ($1 \leq i \leq t$) 为第 i 类买方 Agent 的议题簇集合, 即第 i 类买方 Agent 认为各个议题间存在的依赖关系, r 为议题簇的个数.

定义 6(买方 Agent 类型相同). 买方 Agent 类型集合 $C_A = \{C^1, \dots, C^t\}$ 包含的 t 类买方 Agent 类型是相同的, 当且仅当满足 $C^1 = C^2 = \dots = C^t$.

定义 7(买方 Agent 类型不同). 买方 Agent 类型集合 $C_A = \{C^1, \dots, C^t\}$ 包含的 t 类买方 Agent 类型是不同的, 当且仅当满足 $C^i \cap C^j = \emptyset$, 其中, $1 \leq i \leq t, 1 \leq j \leq t$ 且 $i \neq j$.

定义 8(协商历史矩阵). 协作过滤算法中的 m 行 n 列 User-Item 矩阵 E 定义为协商历史数据矩阵. 设 $U = \{U_1, \dots, U_m\}$ 是矩阵 E 的行集合(也叫用户集合), 其中 U_i 是买方 i 产生的协商历史记录; $I = \{I_1, \dots, I_n\}$ 是矩阵 E 的列集合(也叫议题集合), 其中 I_j 是第 j 个议题. 矩阵 E 是一个二值矩阵, 矩阵中的元素 $e_{ij} = 1$ 表明买方 B 购买了议题 I_j , 否则表示未购买议题 I_j .

定义 9(双向簇). 一个双向簇(biclustor). $A = (U'_b, I'_b) \in 2^{\{1, \dots, m\}} \times 2^{\{1, \dots, n\}}$, 其中用户子集 $U'_b = \{U_1, \dots, U_p\}$ 是矩阵 E 行集合的子集($U'_b \subseteq U$ 且 $p \leq m$), 议题子集 $I'_b = \{I_1, \dots, I_q\}$ 是矩阵 E 列集合的子集($I'_b \subseteq I$ 且 $q \leq n$). 一个双向簇 A 是矩阵 E 的 p 行 q 列子矩阵.

定义 10(包含最大)^[15]. 双向簇 $A = (U'_b, I'_b)$ 是包含最大(inclusion-maximal)的双向簇, 当且仅当 ① $\forall U'_i \in U'_b, I'_j \in I'_b: e_{ij} = 1$ 且 ② $\neg \exists (U''_b, I''_b) \in 2^{\{1, \dots, m\}} \times 2^{\{1, \dots, n\}}$, 其中 (a) $\forall U''_r \in U''_b, I''_s \in I''_b: e_{r,s} = 1$ 且 (b) $U'_b \subseteq U''_b \wedge I'_b \subseteq I''_b \wedge (U'_b, I'_b) \neq (U''_b, I''_b)$.

定义 11(Like 集合). $Like = \{L_1, \dots, L_r\}$ 是关于协商历史矩阵 E 的所有包含最大的双向簇集合, 其中 $\forall L_i = (U'_i, I'_i) \in Like$ ($1 \leq i \leq r$) 都是包含最大

的双向簇.

定义 12(Dislike 集合). 设矩阵 \bar{E} 为协商历史矩阵 E 的反值矩阵. $Dislike = \{D_1, \dots, D_s\}$ 是关于反值矩阵 \bar{E} 的所有包含最大的双向簇集合, 其中 $\forall D_j = (U'_j, I'_j) \in Dislike$ ($1 \leq j \leq s$) 都是包含最大的双向簇.

3 基于效用图的多议题依赖协商模型

3.1 协商模型的形式定义

基于效用图的多议题依赖协商模型可以定义为一个 7 元组 $UGDNM = \langle \Sigma, I, X, C_A, P, U, Pro \rangle$, 其中

(1) $\Sigma = \Sigma_B + \Sigma_S$ 是协商 Agent 的集合. 在一次完整的协商线程中, 本文仅考虑一对一的协商, 并简记某一买方 Agent 为买方 B , 某一卖方 Agent 为卖方 S .

(2) $I = \{I_1, \dots, I_n\}$ 为协商议题(商品)集合, 其中包含有 n 个议题.

(3) $X = \{X_1, \dots, X_n\}$ 是各议题值域集合. $\forall X_i \in X$ ($1 \leq i \leq n$), $X_i = \{0, 1\}$. 即当 x_i 取值为 1 时, 表示购买议题 I_i ; 反之, 表示未购买 I_i .

(4) $C_A = \{C^1, \dots, C^t\}$ 表示买方 Agent 类型集合, 其中 t 为买方 Agent 类型的个数, $C^k = \{C^k_1, \dots, C^k_r\}$ ($1 \leq k \leq t, r = |C^k|$) 为第 k 类买方 Agent 的议题簇集合. 本文在买方 Agent 类型不同的条件下进行研究, 即 $C^i \cap C^j = \emptyset$, 其中, $1 \leq i \leq t, 1 \leq j \leq t$ 且 $i \neq j$.

(5) P 是所有值向量对应的标价构成的集合.

(6) $U = \{U_B, U_S\}$ 是协商 Agent 效用函数组成的集合. 设协商 Agent 采用拟线性效用函数^[11], 即 $U_B(x, p_B) = V_B(x) - p_B(x)$, $U_S(x, p_S) = p_S(x) - V_S(x)$. 设当前参与协商的买方 B 属于第 k 类, 且评价函数 V_B 是一般联加独立效用函数, 即 $V_B(x) = \sum_{i=1}^{|C^k|} u_i(x_{C^k_i})$ ($1 \leq k \leq |C_A|$); 卖方 S 的评价函数 V_S

是联加函数, 即 $V_S(x) = \sum_{i=1}^n Cost(x_i)$, 其中 $Cost(x_i)$ 是每个议题 I_i 的成本, n 为议题的个数. $p_B, p_S \in P$ 分别是买方 B 和卖方 S 对值向量 x 提出的标价.

(7) Pro 表示协商协议, UGDNM 采用轮流提议(alternating-offer)协商协议.

3.2 交互过程

在 UGDNM 的每一步协商中, 协商双方都会产

生一个提议(记为 $O = (x, p)$), 其中 p 是值向量 x 对应的标价.

交互过程可以分为以下两个阶段:

(a) 从集合 C_A 中随机选择某一议题簇集合 C^k 并产生相应的买方 B (即买方 B 属于第 k 类买方 Agent). 买方 B 根据自身的效用图结构及子效用值计算第一个提议(记为 $O_f = (x_f, p_f)$), 并发送给卖方 S . 卖方 S 根据 x_f 的内容提取第 k 类买方 Agent 所产生的协商历史数据, 利用协商历史数据计算各个议题之间的依赖度, 根据各议题间的依赖度学习效用图结构, 并为该效用图分配初始的子效用值.

(b) 卖方 S 利用买方 B 每一次所发送的提议更

新子效用的取值, 并向买方 B 提出使得利润(Gains from Trade, GT)最大的反向提议. 买方 B 对卖方 S 提出的反向提议进行评估并做出相应的让步, 再次发送一个提议给卖方 S , 循环反复直至协商结束. 在协商过程中, 卖方 S 不知道买方 B 的评价函数 V_B 的各子效用取值, 而只是利用一个 Sigmoid 学习函数^[11] 评估各子效用值. 在协商的每一步都使用 2% 的中止协商风险率模拟时间约束.

交互过程如图 2 所示, 其中阶段(b)的过程与文献[11]中描述的内容相同, 本文主要针对阶段(a)的工作进行研究, 即利用协商历史数据学习效用图结构.

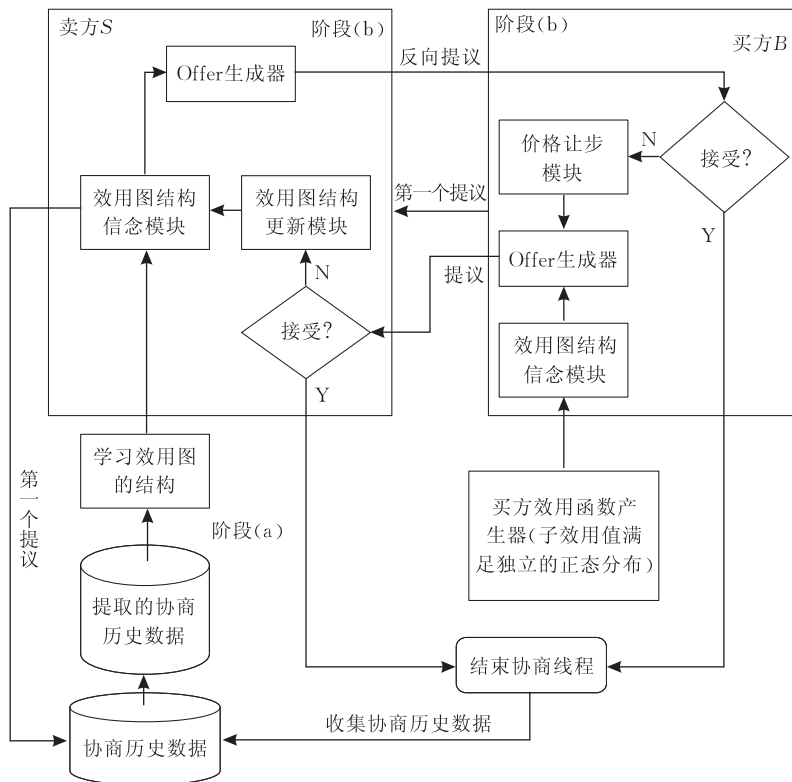


图 2 交互过程

4 基于 Nearest-Biclusters 协作过滤技术的效用图结构学习算法(NBCFL)

本文仅考虑所有议题簇中包含议题个数为 2 的情况, 而关于议题簇中包含更多议题的情况, 将在以后的工作中进行研究.

因为类型相同的买方 Agent 对某些议题簇具有相似的偏好(偏好的局部匹配), 而对于类型不同的买方 Agent 来说, 他们的效用图结构不同, 局部偏好也就不相似. 所以, 当买方 Agent 属于不同类

型时, 利用整个用户集合不能够准确地发现议题间的依赖关系. 本文利用 Nearest-Biclusters 协作过滤技术能够发现买方偏好局部匹配的特性, 提出了卖方 Agent 学习买方 Agent 效用图结构的 NBCFL 算法. NBCFL 算法针对当前参与协商的买方 Agent, 提取与其类型相同的所有买方 Agent 产生的协商历史数据, 然后通过计算议题之间的依赖度学习该类买方 Agent 的效用图结构. NBCFL 算法分为 3 个阶段: (1) 协商历史数据预处理; (2) 求解 Nearest-Biclusters; (3) 学习效用图结构.

4.1 协商历史数据预处理

此阶段采用一种双向聚类算法——Bimax 算法^[15],对协商历史数据进行预处理. Bimax 算法是一种基于分治法的双向聚类算法,该算法能够在二值数据矩阵中较快地(最差情况下的时间复杂度为 $O(mn\beta \min\{m,n\})$,其中 β 是矩阵 \mathbf{E} 中所有包含最大的双向簇的数量, m,n 分别表示矩阵 \mathbf{E} 的行数和列数)找到所有包含最大的双向簇.为了使讨论更加简单,通过一个例子来说明 Bimax 算法.

例 2. Bimax 算法在运行时需要设定一些参数,即设定每个双向簇的最小议题数和最小用户数.该例将两个参数都设为 2(即 $|I'_b| \geq 2$ 和 $|U'_b| \geq 2$),则对协商历史矩阵 \mathbf{E} (如图 3(a)所示)应用 Bimax 算法可以发现 3 个双向簇(如图 3(b)所示):

$$b1: U'_{b1} = \{U_1, U_2\}, I'_{b1} = \{I_1, I_6\};$$

$$b2: U'_{b2} = \{U_1, U_5\}, I'_{b2} = \{I_3, I_5, I_6\};$$

$$b3: U'_{b3} = \{U_4, U_5, U_6\}, I'_{b3} = \{I_2, I_5\}.$$

两个双向簇可能会重叠,即矩阵 \mathbf{E} 中的某些用户或议题可能存在于多个双向簇中.例如, $b1, b2$ 中关于议题 I_6 重叠, $b2, b3$ 关于议题 I_5 重叠.为了不遗漏重要的双向簇,则本文允许重叠的发生,在实验中设定每个双向簇的最小议题数和最小用户数都为 1.

	I_1	I_2	I_3	I_4	I_5	I_6	I_7
U_1	1	0	1	0	1	1	0
U_2	1	0	0	0	0	1	0
U_3	1	0	0	1	0	0	0
U_4	0	1	0	0	1	0	1
U_5	0	1	1	0	1	1	0
U_6	0	1	0	0	1	0	0

(a) 协商历史矩阵 \mathbf{E}

	I_1	I_6	I_3	I_5	I_2	I_7	I_4
U_2	1	1	0	0	0	0	0
U_1	1	1	1	1	1	0	0
U_5	0	1	1	1	1	1	0
U_6	0	0	0	1	1	0	0
U_4	0	0	0	1	1	1	0
U_3	1	0	0	0	0	0	1

(b) Bimax 算法对矩阵 \mathbf{E} 进行双向聚类

图 3

由于 Nearest-Biclusters 协作过滤技术^[13]中讨论的是买方 Agent 最可能购买的议题,而 UGDNM 中议题之间存在着互补性和可替代性的关系,因此本文不仅要考虑买方 Agent 期望购买的议题,还要考虑他们不愿购买的议题.

对于一个二值数据矩阵来说,一个双向簇是矩阵中所有等于 1 的元素构成的子矩阵,所以对协商历史矩阵 \mathbf{E} 的预处理需要做以下两步工作:

(1) 利用 Bimax 算法对协商历史矩阵 \mathbf{E} 中的记录进行分类,将得到的每一个双向簇 $L_i = (U'_i, I'_i)$ 合

并到 Like 集合中.

(2) 求协商历史矩阵 \mathbf{E} 的反值矩阵 $\bar{\mathbf{E}}$. 再次利用 Bimax 算法对数据进行分类,将得到的每一个双向簇 $D_j = (U'_j, I'_j)$ 合并到 Dislike 集合中.

4.2 求解 Nearest-Biclusters

为了学习买方 B 的效用图结构,要将第 k 类买方 Agent 所产生的协商历史数据分离出来.因为买方 B 提出的值向量 \mathbf{x}_f (或反值向量 $\bar{\mathbf{x}}_f$)能够最大程度地反映它的偏好,所以通过计算 Like 集合(或 Dislike 集合)中的双向簇与值向量 \mathbf{x}_f (或反值向量 $\bar{\mathbf{x}}_f$)之间的相似度,找到第 k 类买方 Agent 所产生的协商历史数据.为减少 Bimax 算法中参数设置对相似度计算的影响,本文提出了一种新的相似度计算方法,即值向量 \mathbf{x} 与双向簇 bc 的相似度等于值向量 \mathbf{x} 与双向簇 bc 中共有议题的数量除以二者所含议题数的最大值.

$$\text{sim}(\mathbf{x}, bc) = \frac{|I_x - I_{bc}|}{\max(|I_x|, |I_{bc}|)} \quad (1)$$

其中, I_x 是买方 B 提出的值向量 \mathbf{x} 中期望购买的议题集合, I_{bc} 是双向簇中包含的议题组成的集合,式(1)的值域是 $[0, 1]$. 式(1)表明相似度越高,双向簇中用户子集包含的协商历史数据是由第 k 类买方 Agent 所产生的可能性就越大;反之,则可能性就越小.

算法 1. 寻找最近双向簇.

NearestBiclusters ($\mathbf{Q}, N, \mathbf{x}$)

输入: 双向簇集合 \mathbf{Q} (Like 集合或 Dislike 集合), 阈值 N , 值向量 \mathbf{x}

输出: 行号集合 M

说明: n 是协商议题的个数, nM 是集合 \mathbf{Q} 中的双向簇个数, 数组 $SIM[nM]$ 存放的是集合 \mathbf{Q} 中的双向簇与值向量 \mathbf{x} 的相似度, 变量 nB, nb, cI 分别记录双向簇中包含的议题个数、值向量 \mathbf{x} 中议题取值为 1(即包含的议题)的个数、双向簇和值向量 \mathbf{x} 中共有议题的个数. 集合 M 中的元素是第 k 类买方 Agent 产生的协商历史数据所在行的行号

1. For each $L_s \subseteq \mathbf{Q}$ Do // $0 \leq s \leq nM$, 其中 $L_s = (U'_s, I'_s)$
2. For each $I_i \in I$ Do
3. If $I_i \in I'_s$ and $x_i = 1$ Then $cI \leftarrow cI + 1$; End if
4. If $I_i \in I'_s$ Then $nB \leftarrow nB + 1$; End if
5. If $x_i = 1$ Then $nb \leftarrow nb + 1$; End if
6. End for
7. $SIM[s] \leftarrow cI / \max(nB, nb)$;
8. End for
9. $S \leftarrow \text{Sort}(S, SIM)$; // 根据相似度值将集合 S 中对应的双向簇按降序排列

10. For each $L_s \subseteq S$ and $count < N$ Do

$$//0 \leq s \leq nM, \text{其中 } L_s = (U'_s, I'_s)$$
11. $M \leftarrow U'_s$;
12. If $sim(\mathbf{x}, L_s) \neq sim(\mathbf{x}, L_{s+1})$ Then $count++$;
 End if
13. End for
14. Return (M).

因为值向量 \mathbf{x}_f 受买方 B 所属类型及子效用值的影响,且相似度受到值向量 \mathbf{x}_f 的影响,所以选择相似度大于某一阈值的双向簇作为近邻并不合适.因此,算法 1 对于某一双向簇集合 Q ,选择前 N 个相似度所对应的双向簇作为买方 B 的近邻,并收集这些近邻中包含的用户子集作为提取协商历史记录的依据.

算法 2. 提取协商历史记录.

ExactHistory ($\mathbf{E}, Like, Dislike, N_l, N_d, \mathbf{x}_f, \bar{\mathbf{x}}_f$)

输入:协商历史矩阵 \mathbf{E} , Like 集合, Dislike 集合, 阈值 N_l , 阈值 N_d , 值向量 \mathbf{x}_f , 反值向量 $\bar{\mathbf{x}}_f$

输出:协商历史矩阵 \mathbf{E}'

说明: M_l, M_d, M 是行号集合; SelectRecord 函数的功能是根据集合 M 中的行号从矩阵 \mathbf{E} 中选择协商历史记录存入矩阵 \mathbf{E}' 中

1. $M_l \leftarrow \text{NearestBicluster}(Like, N_l, \mathbf{x}_f)$;
2. $M_d \leftarrow \text{NearestBicluster}(Dislike, N_d, \bar{\mathbf{x}}_f)$;
3. $M \leftarrow M_l \cup M_d$; //将 M_l, M_d 合并为 M ;
4. For each $U_i \in M$ Do
5. $\mathbf{E}' \leftarrow \text{SelectRecord}(\mathbf{E}, U_i)$;
6. End for
7. Return(\mathbf{E}').

算法 2 的目的是利用得到的近邻中包含的用户子集,从矩阵 \mathbf{E} 中提取相应的协商历史记录,合并为一个新的协商历史矩阵 \mathbf{E}' ,并将其作为下一阶段工作的数据集.

4.3 学习效用图结构

经过第 2 个阶段的工作,得到了一个新的协商历史矩阵 \mathbf{E}' . 矩阵 \mathbf{E}' 包含了大量的第 k 类买方 Agent 所产生的协商历史记录.

因为基于关联相似度学习效用图结构的准确率比基于 Cosine 相似度学习的准确率高^[12],所以本文利用基于关联的相似度计算各个议题之间的依赖度.因为基于关联相似度是由基于 Pearson 关联系数(Pearson-Correlation coefficient)演变而来的,而对于二值数据可以将 Pearson 关联系数转化为 ϕ -coefficient^[16].

$Cor(i, j) =$

$$\frac{N_{i,j}(1,1) \times N_{i,j}(0,0) - N_{i,j}(0,1) \times N_{i,j}(1,0)}{\sqrt{N_i(1) \times N_i(0) \times N_j(1) \times N_j(0)}} \quad (2)$$

其中 $N_{i,j}(1,1), N_{i,j}(1,0), N_{i,j}(0,1), N_{i,j}(0,0)$ 是议题 i, j 取值为 0 或 1 时协商记录的个数; $N_i(1), N_i(0), N_j(1), N_j(0)$ 分别是议题 i 或 j 取值为 1 或 0 时协商记录的个数;式(2)的值域是 $[-1, 1]$.

式(2)反映了议题之间的正负关联关系,即依赖度.值越接近于 1 表明两个议题越可能具有较强的互补性,而值越接近于 -1 表明两个议题越可能具有较强可替代性.在理想的情况下,所有存在依赖关系的议题在关联矩阵中的值应该是最大的或最小的.但是受噪声数据的影响,可能会得到一些额外的依赖关系.当然,即使不能够得到 100% 的效用图结构,在交互过程的阶段 (b) (如图 2 所示),卖方 Agent 仍然可以通过买方 Agent 发送的提议来更新子效用值的评估值,从而使协商双方在有限的协商步内找到满足 Pareto 效率的协商结局.

对于卖方 Agent 来说需要解决的问题是:在卖方 Agent 不知道买方 Agent 效用图中实际存在几条依赖边的情况下,应该取多少条依赖边? 一种可能的解决方法是假定效用图中依赖边的个数是图中议题个数的线性函数^[12],即假定效用图中依赖边的个数不会超过 $k_{\max} \times n$ (k_{\max} 是一个较小的正数, n 是协商议题的个数).在本文的实验中,假定效用图中依赖边的个数不超过 $1.5 \times n$ ($k_{\max} = 1.5$).

算法 3. NBCFL 算法.

NBCFL ($\mathbf{E}, \bar{\mathbf{E}}, \mathbf{x}_f, \bar{\mathbf{x}}_f$)

输入:协商历史矩阵 \mathbf{E} , 反值矩阵 $\bar{\mathbf{E}}$, 值向量 \mathbf{x}_f , 反值向量 $\bar{\mathbf{x}}_f$

输出:效用图 G

1. 对矩阵 \mathbf{E} 和 $\bar{\mathbf{E}}$ 调用 Bimax 算法,得到 Like 集合和 Dislike 集合.
2. 调用 ExactHistory ($Like, N_l, \mathbf{x}_f, Dislike, N_d, \bar{\mathbf{x}}_f$), 得到新的协商历史矩阵 \mathbf{E}' .
3. 对于每一个议题对偶,根据式(2)在矩阵 \mathbf{E}' 上计算各个议题间的依赖度,并得到一个 $n \times n$ 的关联矩阵 Cor .
4. 从关联矩阵 Cor 中,分别选择 p 和 q 条(其中 $p+q \leq k_{\max} \times n$) 依赖度取值最大和最小的依赖边,根据选择的依赖边生成效用图 G .

算法 3 的第 4 步中的 p 和 q 条依赖度取值最大和最小的依赖边分别表示议题间存在互补性和可替代性的个数.本文实验中效用图中包含的互补性和可替代性依赖关系的数量比为 2,即 $p/q=2$.

5 实验及分析

实验中协商议题个数为 10,随机产生 3 种类型的买方 Agent.每一类买方 Agent 的效用图结构随

机选取 15 条依赖边, 其中 10 条具有互补性, 5 条具有可替代性. 买方 Agent 对单个议题的评估值及联合效用值随机产生, 并且满足正态分布 $N(1, 0.5)$. 本文采用文献[11-12]的方法收集买方 Agent 发送的值向量 x_f 或协商结局作为协商历史数据. 实验结果取 30 次不同测试结果的平均值. 为了保证这 30 次测试结果尽可能独立, 对于每一次测试都重新生成新的协商历史矩阵, 其中对每次测试都等概率地从 3 类买方 Agent 中选择某一买方 Agent 类型, 并随机生成该类型的 150 个买方 Agent. 对每一个买方 Agent 提出的值向量 x_f 进行检验, 取这 150 个检验结果的平均值作为一次测试结果.

5.1 阈值选择对划分正确率的影响

第 1 组实验(协商历史矩阵 E 包含 1500 条记录)测试阈值选择对划分正确率的影响. 划分正确率, 即对于买方 B 来说, 提取的协商历史矩阵 E' 中由第 k 类买方 Agent 所产生的协商历史的数量与矩阵 E' 中所有历史记录数量的比值. 如图 4 所示, 图中横坐标表示阈值 N_i (或 N_d) 的取值, 纵坐标表示划分正确率. 圆形标示的曲线表示阈值 N_i 对划分正确率的影响, 三角形标示的曲线表示阈值 N_d 对划分正确率的影响. 由图 4 可知, 随着阈值 N_i (或 N_d) 的增加, 所选择的近邻数越来越多, 划分正确率越来越低. 当阈值大于 16 时, 提取的协商历史矩阵 E' 与原协商历史矩阵 E 基本相同(即 $E' = E$), 因此划分正确率为 33.3%. 当阈值为 1 时, 划分正确率最大.

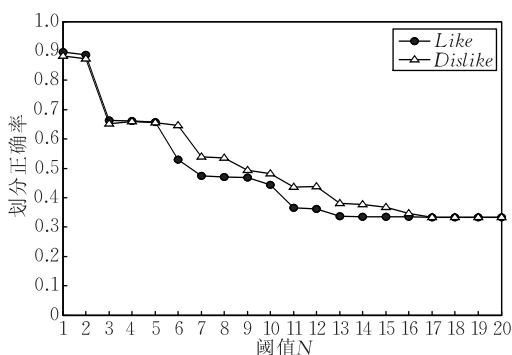


图 4 阈值选择对划分正确率的影响

5.2 阈值选择对效用图结构学习的影响

第 2 组实验(协商历史数据包含 1500 条记录)测试阈值选择对效用图结构学习的影响. 为了说明算法的有效性, 本文采用正确学习到的依赖边个数与买方 Agent 效用图实际依赖边个数的比率作为评估标准, 即效用图结构学习的准确率. 如图 5 所示, 横坐标表示阈值 N_i 或 N_d 的取值, 纵坐标表示效用图结构学习的准确率. 由于 NBCFL 算法需要同

时设定 $Like$ 集合和 $Dislike$ 集合所选近邻的阈值 N_i 和 N_d , 为尽可能地减少两个阈值设定对测试结果的影响, 在测试阈值 N_i 对学习效果的影响时, 设阈值 $N_d = 1$; 同理, 在测试阈值 N_d 对学习效果的影响时, 设阈值 $N_i = 1$. 由图 5 可知, 阈值 N_i 的设定对效用图结构学习准确率的影响最大且当阈值 $N_i = 2$ 和 $N_d = 2$ 时能够获得较高的准确率.

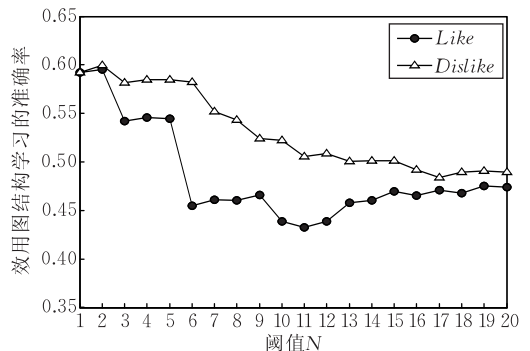


图 5 近邻选择对效用图结构学习准确率的影响

5.3 协商历史矩阵规模对 NBCFL 算法和 IBCFL 算法的影响

最后, 第 3 组实验(阈值 $N_i = 2$ 和 $N_d = 2$)比较分析了当买方 Agent 属于不同类型时协商历史矩阵的规模对 NBCFL 算法与 IBCFL 算法性能的影响. 实验分别对不同规模($m = 300, 600, 900, 1200, 1500, 2700$)的协商历史矩阵进行测试.

由图 6 可知, 在买方属于不同类型的条件下, 使用 NBCFL 算法比 IBCFL 算法能够更加有效地学习效用图的结构. 随着协商历史数据数量的增加, NBCFL 算法学习效用图结构的准确率也随之提高. 由于效用图的结构及子效用值都是随机生成的, 所以不同类型的买方 Agent 可能会提出相同值向量 x_f , 以至于对划分正确率和效用图结构学习的准确率造成负面影响.

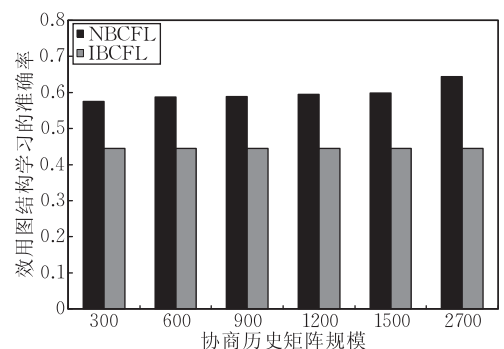


图 6 协商历史矩阵规模对 NBCFL 算法和 IBCFL 算法的影响

通过实验可知,效用图结构学习的准确率受到划分正确率、近邻的选择及协商历史数量等因素的影响.由图 4、图 5 可知,虽然当阈值为 1 时,式(1)能够得到较高的划分正确率,但是此时所得到的协商历史矩阵 E' 中所包含的协商历史记录的数量比较少,而当阈值为 2 时,两方面的影响因素能够取得一个较好的平衡,因而可以得到较好的学习效果.

6 结论与下一步工作

本文基于效用图的多议题依赖协商模型,在参与协商的买方 Agent 类型不同的条件下,对效用图结构学习的问题进行了深入的研究,并基于 Nearest-Biclusters 协作过滤技术的思想提出了卖方 Agent 学习买方 Agent 效用图结构的 NBCFL 算法. NBCFL 算法针对当前参与协商的买方 Agent,提取与其类型相同的所有买方 Agent 产生的协商历史数据,然后通过计算各议题间的依赖度学习该类买方 Agent 的效用图结构.实验表明在参与协商的买方 Agent 类型不同的条件下, NBCFL 算法比 IBCFL 算法能更好地学习买方 Agent 的效用图结构.

今后的工作将主要集中在进一步提高效用图结构学习的准确率以及在子簇内包含有更多议题情况下的依赖性研究.

参 考 文 献

- [1] Faratin P. Automated service negotiation between autonomous computational agents [Ph. D. dissertation]. Department of Electronic Engineer Queen Mary & Westfield College, University of London, London, 2000
- [2] Coehoorn R M, Jennings N R. Learning an opponent's preferences to make effective multi-issue negotiation tradeoffs//Proceedings of the 6th International Conference on E-Commerce. Delft, 2004; 59-68
- [3] Wang Li-Ming, Huang Hou-Kuan, Chai Yu-Mei. Choosing multi-issue negotiating object based on trust and K-armed bandit problem. Journal of Software, 2006, 17(12): 2537-2546(in Chinese)
(王黎明, 黄厚宽, 柴玉梅. 基于信任和 K 臂赌博机问题选择多问题协商对象. 软件学报, 2006, 17(12): 2537-2546)
- [4] Wang Li-Ming, Huang Hou-Kuan. A multistage-based framework for multi-Agent multi-issue negotiation. Journal of Computer Research & Development, 2005, 42(11): 1849-1855(in Chinese)
(王黎明, 黄厚宽. 一个基于多阶段的多 Agent 多问题协商
- 框架. 计算机研究与发展, 2005, 42(11): 1849-1855)
- [5] Wang Li-Ming, Huang Hou-Kuan. Speculative computation based on master Agent belief revision and its resource negotiation. Journal of Software, 2005, 16(11): 1920-1928(in Chinese)
(王黎明, 黄厚宽. 基于主 Agent 信念修正的推测计算及其资源协商. 软件学报, 2005, 16(11): 1920-1928)
- [6] Klein M, Faratin P, Sayama H et al. Negotiating complex contracts. Group Decision and Negotiation, 2003, 12(2): 111-125
- [7] Takayuki Ito, Klein Mark, Hattori, Hironitsu. An auction-based negotiation protocol for agents with nonlinear utility functions. MIT Sloan School of Management, Cambridge, Massachusetts, USA, Research Paper; No. 4597-06, 2006
- [8] Takayuki Ito, Klein Mark. A multi-issue negotiation protocol among competitive agents and its extension to a nonlinear utility negotiation protocol//Proceedings of the 5th International Joint Conference on Autonomous Agents and Multi-Agent Systems(AAMAS06). New York, USA: ACM Press, 2006; 435-437
- [9] Lai Guoming, Li Cuihong, Sycara Katia. Efficient multi-attribute negotiation with incomplete information. Group Decision and Negotiation, 2006, 15(5): 511-528
- [10] Somefun Koye, Klos Tomas B, Poutré Han La. Online learning of aggregate knowledge about nonlinear preferences applied to negotiating prices and bundles//Proceedings of the 6th International Conference on Electronic Commerce (ICEC'04). New York, USA: ACM Press, 2004; 361-370
- [11] Robu V, Somefun K, Poutré Han La. Modeling complex multi-issue negotiations using utility graphs//Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems. New York, USA: ACM Press, 2005; 280-287
- [12] Robu V, Poutré Han La. Learning the structure of utility graphs used in negotiation through collaborative filtering//Proceedings of the 8th Pacific Rim Workshop on Multi-Agent Systems (PRIMA'05), K. L., Malaysia, 2005; 192-209
- [13] Symeonidis P, Nanopoulos A, Papadopoulos A N, Manolopoulos Y. Nearest- biclusters collaborative filtering//Proceedings of the 8th ACM SIGKDD Workshop on Web Mining and Web Usage Analysis (WEBKDD 2006). Philadelphia, Pennsylvania, 2006; 36-55
- [14] Gonzales C, Perny P. GAI Networks for utility elicitation//Proceedings of the 9th International Conference on the Principles of Knowledge Representation and Reasoning. Whistler, Canada, 2004; 224-234
- [15] Amela Prelic et al. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics, 2006, 22(9): 1122-1129
- [16] Tan Pang-Ning, Steinbach Michael, Kumar Vipin. Introduction to Data Mining. New Jersey: Addison-Wesley, 2006



WANG Li-Ming, born in 1963, Ph. D., professor. His main research interests are distributed artificial intelligence and data mining.

LI Kun, born in 1982, M. S.. He is mainly engaged in the research of distributed artificial intelligence.

Background

Multi-issue negotiation generally assumes that negotiation issues are independent, but real preferences often exhibit dependencies among issues. Graphical utility models represent powerful formalisms for modeling agents' non-linear preferences. The multi-issue negotiation model based on utility graph which considers interdependencies between issues enables agents to reach Pareto-efficient agreements with a limited number of negotiation steps. The key problem of the negotiation model is how to learn the structure of utility graph effectively. Unlike IBCFL algorithm only considering the condition that buyers from the same class of buyers,

NBCFL algorithm in this paper tackles the problem of learning the structure of Utility Graph on condition that buyers from different classes of buyers. Firstly, the algorithm takes advantage of the trait that nearest-biclusters collaborative filtering could detect partial matching of buyers' preferences, and collects the negotiation data which were produced by buyer that is the same class with active buyer. Secondly, it retrieves the structure of utility graph using the degree of interdependencies between issues. The purpose of these steps is to learn the structure of utility graph more effectively.