ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.12, December 2006, pp.2537−2546
DOI: 10.1360/jos172537

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于信任和 K 臂赌博机问题选择多问题协商对象*

王黎明[1+], 黄厚宽[2], 柴玉梅[1]

[1](郑州大学 信息工程学院,河南 郑州 450052)

[2](北京交通大学 计算机与信息技术学院,北京 100044)

Choosing Multi-Issue Negotiating Object Based on Trust and K-Armed Bandit Problem

WANG Li-Ming[1+], HUANG Hou-Kuan[2], CHAI Yu-Mei[1]

[1](School of Information Engineering, Zhengzhou University, Zhengzhou 450052, China)

[2](School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

+ Corresponding author: Phn: +86-371-67762865, Fax: +86-371-67761542, E-mail: cymwlm@zzu.edu.cn, http://www.zzu.edu.cn

Wang LM, Huang HK, Chai YM. Choosing multi-issue negotiating object based on trust and K-armed bandit problem. *Journal of Software*, 2006,17(12):2537−2546. http://www.jos.org.cn/1000-9825/17/2537.htm

Abstract: Multi-Issue negotiation between Agents is a complicated course in which negotiating Agents mutually exchange offers. Solving the problem of choosing seller before negotiation has important practical value in e-commerce. The problem is solved in this paper to improve accuracy of the multi-issue negotiation and buying Agent's utility. In order to fully utilize negotiation history, tradeoff exploration and exploitation, the problem of choosing seller is transformed into a K-armed bandit problem. A model for measuring trust and reputation is presented, several improved algorithms, which are used to learn reward distribution and combine learning with technologies for K-armed bandit problem, are presented. Finally, the combination of the improved algorithms, the trust and reputation improves the accuracy and practicability of choosing a selling Agent. Several experiments prove validity of the work in application.

Key words: Agent; negotiation; K-armed bandit problem; trust; reputation; utility

摘 要: Agent 之间的多问题协商(multi-issue negotiation)是一个复杂的动态交互过程.解决协商之前的对象选择问题在电子商务中有着重要的应用价值.为了提高多问题协商的准确性和购物 Agent 的效用,主要解决协商前的销售 Agent 的选择问题.为了充分利用协商历史,实现探索(exploration)和利用(exploitation)的折衷,把销售 Agent 的选择问题转变成 K 臂赌博机问题(K-armed bandit problem)来求解.提出了信任和声誉的度量模型,结合 K 臂赌博机问题的求解技术,采用学习机制,提出了几个确定奖励分布的改进算法.最后,以模拟协商过程为基础,将改进算法、信任和声誉有机地结合起来,提高了选择销售 Agent 的准确性和实用性.几个实验都说明了该工作在应用中的有效性.

关键词: Agent;协商;K 臂赌博机问题;信任;声誉;效用

中图法分类号: TP18 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant No.60443003 (国家自然科学基金)
Received 2005-01-12; Accepted 2006-03-07

在基于 Web 的电子市场中,当一个购物 Agent(购买者)带着用户的需求来到市场时,它将面对多个销售 Agent(销售者).购买者需要按照用户的需求选择合适的销售者,以使自己的协商结果效用最大.文献[1−5]分别给出了不同策略下的多问题协商模型,但都没有考虑协商前协商对象的选择问题.协商之前解决这个问题在现实中具有实际的意义.购买者只能利用对销售者的信任程度、其他购买者的建议以及相应的协商历史进行选择.文献[6]给出了一个信任模型:信任+不信任+不确定性=1,该模型基于"建议"这个四元组概念对信任进行度量.文献[7]列举了信任和不信任的特性,并对这些特性进行了分析.文献[8]认为,如果销售者的可信任程度等于购买者对他的信任程度,那么社会福利(social welfare)、交易量和协商 Agent 的效用都可以达到最大.本文将销售者选择问题转变成 K 臂赌博机问题来求解.文献[9]为求解 K 臂赌博机问题提出了 UCB1,UCB2,ε_n-GREDDY 和 UCB1-NORMAL 这 4 种策略.文献[10]借助于 K 臂赌博机问题求解技术,提出一种 *Hedge*(β)算法以解决资源的动态分配问题.

本文提出了信任-声誉(trust and reputation)模型,结合 K 臂赌博机问题求解技术和学习机制,并通过模拟协商的方式建立购买者对销售者的信任度和声誉度.购买者利用 K 臂赌博机问题中的无假设技术,动态产生各个销售者的奖励分布.以 *Hedge*(β)算法为基础给出了几种改进算法,最后,购买者将信任、声誉与奖励最大化地结合起来,选择使自己收益最大的销售者.

1 多问题协商的相关定义

定义 **1**(协商 **Agent** 集合). $\Sigma=\Sigma_B+\Sigma_S$ 表示参与协商的 Agent 集合,其中Σ_B 表示购买者集合,Σ_S 表示销售者集合.

定义 **2**(多问题向量及其值向量). $Iss=\{i_1,i_2,\ldots,i_n\}$表示协商过程中涉及到的 n 个问题,其中 n 表示协商涉及的问题数.那么, $\overrightarrow{Iss}=\langle i_1,i_2,\ldots,i_n\rangle$表示协商过程中 n 个问题的问题向量,其中$\forall m,1\le m\le n,i_m\in Iss$ 表示第 m 个协商问题. $\vec{V}^{\overrightarrow{Iss}}=\langle v^{i_1},v^{i_2},\ldots,v^{i_n}\rangle$ 表示在多问题向量 $\overrightarrow{Iss}$ 上的一个取值向量,其中$\forall i_m\in\overrightarrow{Iss}$, $v^{i_m}\in\Omega_m$ 表示问题 i_m 的一个取值,Ω_m 表示问题 i_m 取值的值域.

定义 **3**(单问题效用及多问题联合效用). $\forall a\in\Sigma,\forall i_m\in\overrightarrow{Iss}$, $\xi_a^{i_m}$ 表示 Agent a 在问题 i_m 上的效用函数,是 $v_a^{i_m}$ 到实值空间的映射,即 $\xi_a^{i_m}:\{v_a^{i_m}\}\to R$,其中, $v_a^{i_m}\in\Omega_m$ 表示 Agent a 在问题 i_m 上的一个取值.JU_a:$\Omega_1\times\Omega_2\times\ldots\times\Omega_n\to R$ 是 Agent a 在多问题向量 $\overrightarrow{Iss}$ 上的效用函数.对于 Agent a 和多问题向量 $\overrightarrow{Iss}$, $JU_a(\vec{V}^{\overrightarrow{Iss}})$ 是其多个问题效用值的加权和,即

$$JU_a(\vec{V}^{\overrightarrow{Iss}})=\sum_{k=1}^{n}w_a^{i_k}\times\xi_a^{i_k}(v^{i_k}).$$

其中,$\forall i_k\in\overrightarrow{Iss}$, $w_a^{i_k}$ 表示 $JU_a(\vec{V}^{\overrightarrow{Iss}})$ 中问题 i_k 的效用函数 $\xi_a^{i_k}(v^{i_k})$ 的权值,且 $\sum_{k=1}^{n}w_a^{i_k}=1$.

更完整的多问题协商模型请参考文献[1−5,12,13].

2 信任-声誉模型

购买者需要利用信任和声誉在销售者中选择自己信任并能给自己带来较高效用的销售者.信任是购买者对销售者在可靠性、诚实度和能力方面的信念,而声誉是购买者从其他购买者那里获得的关于销售者的信念.例如,Agent a 和 Agent b 有过交易,而且交易给 a 带来较好的效用,那么 a 对 b 就有信任关系.如果 a 和 b 没有交易过,那么 a 可以通过 b 的声誉获得信任.b 的声誉就是 a 通过其他 Agent 获得的关于 b 的信息,并且这些信息对 a 有益,从而 a 对 b 有比较好的信任关系.信任和声誉都是一种信念.

2.1 问题的提出

设购买者集合$\Sigma_B=\{b_1,b_2,\ldots,b_n\}$和销售者集合$\Sigma_S=\{s_1,s_2,\ldots,s_m\}$,购买者 $b_i\in\Sigma_B$ 要从某个 $s_j\in\Sigma_S$ 那里购买产品,问题是购买者 b_i 如何从Σ_S 中选择一个 s_j 以使其效用最大.为此,我们给出如下相关定义:

定义 **4**(信任依赖图). $TrustG=\langle V,E,D\rangle$.

- V 表示顶点集合,$V=V_1+V_2$,V_1 为边的发出顶点集合,V_2 为边的接收顶点集合.
- E 表示有向边集合,$E=E_1+E_2$,E_1 为实线有向边集合,一条实线有向边从 $v_i\in V_1$ 发出到达 $v_j\in V_2$,它表示 v_i 对 v_j 有信任关系,记为(v_i,v_j);E_2 为虚线有向边集合,一条虚线有向边从 $v_k\in V_1$ 到达 $v_f\in V_1$,它表示 v_k 对 v_f 的推荐有信任关系,记为 $\overline{(v_k,v_f)}$.
- D 是标示在实(虚)线有向边上的实数集合,实数表示信任度量或者声誉度量.如果 $v_i\in V_1,v_j\in V_2$ 且 $(v_i,v_j)\in E_1$,则 $Tr^{(v_i,v_j)}\in D$ 表示 v_i 对 v_j 的信任度.

定义 **5**(**Agent *b*** 对 **Agent *s*** 有信任关系). G 是一个 $TrustG,(b,s)\in E_1,b\in V_1,s\in V_2$.如果 b 和 s 之间有过 n 次交易,$n\geq 1$,并且这 n 次交易所获得的效用分别是 $u_1,u_2,\ldots,u_n$,最满意效用分别是 $u_1^*,u_2^*,\ldots,u_n^*$,那么,称 b 对 s 有信任关系(直接信任关系),记为 $T_{b\to s}$.对这种信任关系的度量定义如下:

$$T_{b\to s}(u_1,u_2,\ldots,u_n,u_1^*,u_2^*,\ldots,u_n^*)=\sum_{i=1}^{n}u_i/\sum_{i=1}^{n}u_i^*.$$

Agent b 有两个阈值 θ_1,θ_2,且 $0\leq\theta_1\leq\theta_2$.如果 $T_{b\to s}\geq\theta_2$,则 Agent b 信任 Agent s;如果 $\theta_1\leq T_{b\to s}\leq\theta_2$,Agent b 弱信任 Agent s;否则,Agent b 不信任 Agent s.

定义 **6**(**Agent *s*** 的声誉). G 是一个 $TrustG,(b,s)\notin E_1,b\in V_1,s\in V_2$.除 b 以外的所有购买者对 s 的信任综合称为相对于 b 的 s 的声誉,记为 $R_{b\to s}$.对声誉的度量定义如下:

$$R_{b\to s}(r_1^b,r_2^b,\ldots,r_{|V_1|-1}^b)=\sum_{i=1,(b_i,s)\in E_1\text{且}(b,s)\notin E_1}^{|V_1|-1}r_i^bT_{b_i\to s}.$$

其中,r_i^b 表示 b 对每个 b_i 的信任程度,且 $\sum_{i=1}^{|V_1|-1}r_i^b=1$.Agent b 有两个阈值 θ_1,θ_2,且 $0\leq\theta_1\leq\theta_2$.如果 $R_{b\to s}\geq\theta_2$,则 Agent b 认为 Agent s 是可信任的;如果 $\theta_1\leq R_{b\to s}\leq\theta_2$,Agent b 认为 Agent s 是弱可信任的;否则,Agent s 是不可信任的.

2.2 选择协商对象的过程

在 Σ 中考虑购买者和销售者.从购买者的角度实施选择销售者的过程.一个购买者 b_i 在选择一个销售者 s_k 时可能出现两种情况:一种情况是 b_i 和 s_k 有过交易,那么,b_i 是否选择 s_k 就取决于它对 s_k 的信任加上 s_k 的声誉;另一种情况是 b_i 和 s_k 没有交易过,那么,b_i 是否选择 s_k 就只能取决于 s_k 的声誉.

协商对象选择过程:Ch_P.

说明:该过程首先是阅读多个购买者和销售者之间的协商历史,识别交易关系,通过历史计算他们之间的信任度和声誉度,确定购买者和销售者之间的信任依赖图,然后根据信任依赖图选择销售者.

1. 初始化

购买者顶点集 $V_1\leftarrow\varnothing$,销售者顶点集 $V_2\leftarrow\varnothing$,实线边集 $E_1\leftarrow\varnothing$,虚线边集 $E_2\leftarrow\varnothing$, $D\leftarrow\varnothing$.

2. 识别交易关系

对 $\forall b\in\Sigma_B$,扫描 $\forall s\in\Sigma_S$,如果 b 和 s 有交易关系,则 $V_1\leftarrow V_1\cup\{b\},V_2\leftarrow V_2\cup\{s\},E_1\leftarrow E_1\cup\{(b,s)\}$.

3. 计算直接信任度

对 $\forall b\in V_1,\forall s\in V_2$,如果 $(b,s)\in E_1$,则识别 b 和 s 的交易历史,并计算 $T_{b\to s}$,即 $Tr^{(b,s)}=T_{b\to s},D\leftarrow D\cup\{T_{b\to s}\}$.对于 $\forall b_i,b_j\in V_1$,识别 $Tr^{\overline{(b_i,b_j)}}$ 和 $Tr^{\overline{(b_j,b_i)}}$,并且 $D\leftarrow D\cup\{Tr^{\overline{(b_i,b_j)}}\}$, $D\leftarrow D\cup\{Tr^{\overline{(b_j,b_i)}}\}$.

4. 对每个购买者计算声誉度

为每个购买者 $b\in V_1$ 建立一个声誉度表 RE_b.对 $\forall b\in V_1,\forall s\in V_2$,如果 $(b,s)\notin E_1$,则计算 $R_{b\to s},RE_b\leftarrow RE_b\cup\{R_{b\to s}\}$.

5. 按照策略选择销售者

有两种选择策略:一种是在曾经交易过的对象中选择.此时,购买者依靠信任依赖图中的直接信任度和声誉度进行选择,即购买者 b 将选择使 $T_{b\to s}+R_{b\to s}$ 最大的销售者,S^* 表示这个销售者,

$$S^*=\arg\max_{s\in V_2\text{且}R_{b\to s}\in RE_b}(T_{b\to s}+R_{b\to s}).$$

另一种是在未曾交易过的对象中选择,此时,购买者只依靠信任依赖图中的声誉度来选择,即购买者 b 将选

择使 $R_{b\to s}$ 最大的销售者,

$$S^* = \underset{s\in V_2,(b,s)\notin E且R_{b\to s}\in RE_b}{\arg\max}(R_{b\to s}).$$

后一种策略有助于开拓新的交易对象,对购买者以后购买产品有益.对于具有长期购买任务的购买者来说,对小规模购买任务选择后者是有意义的.

3 协商对象选择问题到 K 臂赌博机问题的转换

3.1 K臂赌博机问题

在原始的 K 臂赌博机问题中,有一个拥有 K 个手臂(arms)的投币机器(slot machine),赌博者要从这 K 个手臂中选择一个手臂进行操作来获得奖励(reward),这个奖励可能是正值、0 或者负值,一次操作(play)获得一个奖励.在一个特定的时间段内,赌博者只能操作一个手臂,并且使赌博者的奖励总和最大.假定每个手臂的奖励有不同的分布,那么,赌博者要尽快找到使自己获得最大奖励的手臂,并用这个手臂进行赌博.

3.2 协商对象选择问题

对于$\forall b\in\Sigma_B,|\Sigma_S|=K$(相当于 K 臂赌博机问题中的手臂数),$b$(相当于 K 臂赌博机问题中的赌博者)从$\Sigma_S$中选择一个销售者(相当于 K 臂赌博机问题中的手臂)进行协商.在 b 对Σ_S中的销售者不了解的情况下,它需要从以前的协商历史中学习,获得来自不同销售者的奖励分布.$\forall s_i\in\Sigma_S,i=1,2,\ldots,K$ 都对应一个协商历史 $H^{s_i}=\langle HT_1^{s_i},HT_2^{s_i},\ldots,HT_m^{s_i}\rangle$,其中,$m$ 表示协商历史的长度; $HT_i^{s_i}$ 表示历史中的一个线程.b 通过阅读协商历史模拟和 s_i 的协商过程,从中获得奖励的分布情况.b 阅读一个 $HT_i^{s_i}$,相当于选择 s_i 执行了一次操作.

对于 $HT_j^{s_i}=\langle O_{j,1}^{s_i},O_{j,2}^{b'},\ldots,O_{j,y}^{s_i}\rangle$,其中,$y$ 是线程长度,假设每个线程中的第 1 个元素 $O_{j,1}^{s_i}$ 由销售者 s_i 首先给出.如果 b 面对历史中 s_i 的当前 offer 是 $O_{j,2k+1}^{s_i}$,且 b 给出的一个模拟反 offer 表示为 O_k^b,历史中对应的反 offer 为 $O_{i,2k+2}^{b'}$,那么,$\Delta_k^{s_i}=JU_b(O_{i,2k+2}^{b'})-JU_b(O_k^b)$ 表示一次交互的奖励.如果 b 的第 n 次操作选择的是 s_i,那么,执行本次操作所获得的奖励为

$$r^{s_i}(n)=\sum_{k=1}^{\lfloor y/2\rfloor}\Delta_k^{s_i}.$$

假设算法 A 阅读协商历史获得奖励,并探索下一个要模拟协商的对象. $F^{s_i}(N)$ 表示执行 N 次操作中选择 s_i 的次数,执行 N 次操作后,该算法的遗憾(regret)是

$$R_\Delta^A=r^*N-\max_{1\le i\le K}\left(r^{s_i}\sum_{i=1}^{K}IE[F^{s_i}(N)]\right),$$

其中 $r^*\overset{\text{def}}{=}\max_{1\le i\le K}(r^{s_i})$,$IE[\cdot]$表示期望.$ER_{best}$ 表示在 N 次操作之后,最好动作所产生的总奖励,

$$ER_{best}=\max_{1\le i\le K}\left(r^{s_i}\sum_{i=1}^{K}IE[F^{s_i}(N)]\right).$$

R_Δ^A 表示算法 A 在 N 次操作中奖励的损失.在 N 次操作中选择最优协商对象的次数越多,奖励损失就越少,算法性能就越优.算法在$\{s_1,s_2,\ldots,s_K\}^N$上确定 N 次操作的奖励概率分布, $r^{s_i}(n)$ 是在$\{s_1,s_2,\ldots,s_K\}^{n-1}$上选择第 n 个对象是 s_i 的随机变量.

3.3 模拟协商及奖励分布估计算法

3.3.1 模拟协商算法

模拟协商是购买者 $b\in\Sigma_B$ 阅读满足其购物要求的协商历史的过程.在协商过程中,b 阅读销售者的 offer 和原购买者的 offer,并根据销售者的 offer 模拟产生自己的 offer,将原购买者的 offer 与自己 offer 的效用之差作为一次交互的奖励.

模拟协商算法:$N(s_i)$.

算法执行一次,购买者 b 和所选销售者 s_i 完成一个线程的模拟协商,将协商后的奖励之和作为一次选择操作的奖励.假设当前阅读的成功线程是 s_i 的 $HT_i^{s_i}$,并且一开始有一个指针指向 $HT_i^{s_i}$ 的第 1 个元素,那么每读一个 offer,指针自动后移一个位置.指针所指元素表示为 $HT_i^{s_i}\uparrow$.

(1) $O^{s_i}\leftarrow HT_i^{s_i}\uparrow$,如果销售者的 O^{s_i} 是结束 offer,则模拟协商结束,返回奖励和.

(2) $O^{b'}\leftarrow HT_i^{s_i}\uparrow$,如果原购买者的 $O^{b'}$ 是结束 offer,则模拟协商结束,返回奖励和.

(3) 购买者 b 按照 O^{s_i} 产生自己的 offer 为 O^b. $\Delta^{s_i}=JU_b(O^{b'})-JU_b(O^b)$,如果 $\Delta^{s_i}>0$ 则是正奖励;否则是无奖励,即

$$r^{s_i}\leftarrow\begin{cases}r^{s_i}+\Delta^{s_i}, & \text{如果}\Delta^{s_i}>0\\ r^{s_i}, & \text{否则}\end{cases}.$$

(4) 如果 $HT_i^{s_i}\uparrow$ 为空,则结束模拟协商,返回奖励和;否则转(1).

3.3.2 奖励分布估计算法

通常,求解 K 臂赌博机问题时一般都对每个手臂所产生的奖励给出一些统计的假设,比如,每个手臂产生奖励的分布被假设成是高斯和时间不变的.但实际上,这种假设可能是不合适的,或者很难给出合适的假设,甚至不可能给出合适的假设.文献[11]在求解 K 臂赌博机问题时,对每个手臂采用了无统计假设技术,在 T 个实验时间步中,产生手臂所获得的奖励分布.

本文以文献[10]中的 $Hedge(\beta)$算法为基础提出几种改进算法,并以模拟协商为基础,将 K 臂赌博机问题求解技术[11]应用到各种算法中去,从而产生选择各个协商对象所获得的奖励分布.

分布估计算法:$\boldsymbol{H(\alpha,n)}$.

本算法以$\alpha>0$ 为参数确定选择协商对象 s_i 的概率分布.本算法只考虑了第 n 步模拟协商时选择协商对象的分布.$RS_i(n)$表示前 $n-1$ 步选择 s_i 的奖励累计.

初始化:

for $i=1,\ldots,K$ do $RS_i(1)\leftarrow 0$.

for $i=1,\ldots,K$ do

{1. 产生概率分布向量 $P(n)$,选择 s_i 的概率为 $p_i(n)$

$$p_i(n)=(1+\alpha)^{RS_i(n)}\Big/\sum_{j=1}^{K}(1+\alpha)^{RS_j(n)}$$

2. 执行算法 $N(s_i)$和 s_i 进行模拟协商,获得 $r^{s_i}(n)$, $r^{s_i}(n)\leftarrow V_s(r^{s_i}(n))$, $V_s(\cdot)$是一个奖励评分函数,且 $r^{s_i}(n)\in[0,1]$,它是奖励向量 $R(n)$中的第 i 个元素.

3. 产生每个 s_i 的累计奖励:

$$RS_i(n+1)\leftarrow RS_i(n)+r^{s_i}(n).$$

}

算法 $H(\alpha,n)$只考虑在一步中各个协商对象被选择的可能性,模拟协商要扫描更多的协商线程才可以获得较精确的分布信息.如果执行 for $n=1,\ldots,N$ do $\{H(\alpha,n)\}$就可以获得各个协商对象在 N 步之内的概率分布.这种算法的主要目的是为选择协商对象产生一个概率分布.按照这个分布以及奖励可以选择协商对象:

$$S^*=\arg\max_{1\le i\le K}\left(\sum_{j=1}^{K}p_i(j)r^{s_i}(j)\right).$$

其中,S^*表示在这个分布下能获得最大奖励的协商对象,该算法的复杂度为 $O(N^2)$.但是,这种方式产生的分布可能使协商对象在奖励方面差距较大,比如,有可能具有较大奖励的对象却有较低的选择概率.为此,我们将算法产生的分布 $P(n)$和均匀分布叠加产生新的分布,以保证算法尝试到所有对象,并为每个对象得到准确的奖励估计.为此,我们给出算法 $E(\alpha,\gamma)$,算法复杂度依然是 $O(N^2)$.

分布叠加算法:$E(\alpha,\gamma)$.

本算法调用算法 $H(\alpha,n)$,然后将分布 $P(n)$和均匀分布叠加产生新的分布 $\hat{P}(n)$.本算法以α和γ为参数,其中$\alpha>0$ 用于 $H(\alpha,n)$,$\gamma\in[0,1]$用于分布叠加.

for n=1,…,N do

{ 调用 $H(\alpha,n)$获得分布 $P(n)$.

for i=1,…,K do { $\hat{p}_i(n)\leftarrow(1-\gamma)p_i(n)+\gamma/K$. $\hat{r}^{s_i}(n)\leftarrow(\gamma/K)\times(r^{s_i}(n)/\hat{p}_i(n))$.}

}

对于被选择的 s_i,叠加奖励 $\hat{r}^{s_i}(n)$ 和 $r^{s_i}(n)/\hat{p}_i(n)$ 成正比例关系,并且,其期望叠加奖励也正比例于其实际奖励,即 $E^{s_n}[\hat{r}^{s_i}(n)\mid s_1,...,s_{n-1}]=\frac{\gamma}{K}r^{s_i}(n)$,其中,对于 $1\le j\le n$,$s_j\in\Sigma_S$,并且$\langle s_1,\ldots,s_n\rangle$是$\{s_1,\ldots,s_K\}^n$ 上的一个长度为 n 的销售者序列;因子γ/K 是为了保证 $\hat{r}^{s_i}(n)\in[0,1]$.

3.4 模拟参数α和γ的选择

在算法 $H(\alpha,n)$中,参数α是概率分布 $P(n)$的制约参数,而参数γ是算法 $E(\alpha,\gamma)$进行分布叠加的制约参数.参考文献[11]中的引理 4.2 调整了这两个参数,并给出如下引理:

引理. 如果 $f\ge ER_{best}$,并且算法 $E(\alpha,\gamma)$中的α和γ分别选取 $\alpha=\sqrt[3]{(4K\ln K)/f}$, $\gamma=\min\{1,\sqrt[3]{(K\ln K)/(2f)}\}$,那么,算法 $E(\alpha,\gamma)$的遗憾满足

$$R_A^{E(\alpha,\gamma)}\le\frac{3}{\sqrt[3]{2}}f^{2/3}(K\ln K)^{1/3}.$$

从引理看出,可以动态地寻找关系 $f\ge ER_{best}$.满足此关系并维持上述遗憾关系的α和γ称为稳定点.在寻找过程中,要时刻记录与每个销售者进行模拟协商后所得到的奖励,用最大值和 f 进行比较,f 可以是任何一个递增的量,本文在这里选择了 $f\leftarrow 2^n$, $\hat{ss}_i$ 表示与 s_i 进行模拟协商所获得的奖励和.实现这个动态过程的算法是 DS.

确定参数α和γ算法:DS.

说明:动态寻找关系 $f\ge ER_{best}$,确定参数α和γ.

初始化:$n\leftarrow 1$.

(1) $f\leftarrow 2^n$. $\alpha\leftarrow\sqrt[3]{(4K\ln K)/f}$. $\gamma\leftarrow\min(1,\sqrt[3]{(K\ln K)/(2f)})$.执行算法 $E(\alpha,\gamma)$. for i=1,…,K do $\hat{ss}_i\leftarrow 0$. $k\leftarrow 1$.

(2) for i=1,…,K do $\hat{ss}_i\leftarrow\hat{ss}_i+r^{s_i}(k)/\hat{p}_i(k)$.

(3) If $f<\max_i(\hat{ss}_i)+K/\gamma$ then $n\leftarrow n+1$,转(1). else if $k=N$ then 算法结束.$k\leftarrow k+1$,转(2).

如果$f<\max_i(\hat{ss}_i)+K/\gamma$,说明还不满足引理中的条件,需要继续通过 n 增加 f,调整α和γ并重新执行算法 $E(\alpha,\gamma)$,产生各个 s_i 的奖励及其分布;如果 $f\ge\max_i(\hat{ss}_i)+K/\gamma$,并且 $k<N$,那么,对每一步的各个 s_i 所得奖励继续进行累计.N 值越大,奖励的分布就越精确,由最大奖励和所决定的 s_i 就越准确. $\hat{ss}_i$ 中累加 $r^{s_i}/\hat{p}_i(k)$,其中, $\hat{p}_i(k)$ 表示购买者选择对象 s_i 的概率,除以 $\hat{p}_i(k)$ 是为了保证每个估计的期望值是正确的,也就是说,对于 $1\le i\le K$ 和 $1\le k\le N$,有

$$E[\hat{ss}_i(k)]=\sum_{k'=1}^{k}r^{s_i}(k').$$

3.5 信任和声誉与奖励最大化的结合

购买者通过模拟协商探索最好的协商对象,在探索过程中,累计每个可能对象所提供的奖励,N 步之后,可以选择获得累计奖励最大的协商对象.但是,在此基础上,还要考虑这样几个因素:(1) 销售者的协商历史可能是不丰富的,即他可能是一个新手,或者是一个新来者;(2) N 可能不足够大,从而影响上述方法的精确度;(3) 协商历史与现在的时间距离会影响累计奖励对协商对象的选择;(4) 外来因素也是决定选择协商对象的重要因素之一.如果因素(1)出现,上述方法的结果就不足以选择最好的协商对象,自然,N 也不可能很大,上述方法的精度可

能会不足,这时,需要其他信息来补充.因素(3)说明,购买者学习的历史越久远,学习结果对选择协商对象的支持力度越小.因素(4)是决定选择对象的另一个方面的内容,是前 3 个因素的补充.为了避免前 3 个因素出现降低选择精度,本文将上述方法与因素(4)结合起来.因素(4)的信息由信任度、声誉度及时间因素组成,信任度和声誉度通过学习或者主观获得.为了解决上述相关因素所产生的问题,下面引入几个相关向量的定义.

定义 **7**(时间折扣因子向量). $\delta^{b_i}=\langle\delta_1,\delta_2,...,\delta_n\rangle$ 表示时间折扣因子向量.其中,δ_j 表示 $b_i\in\Sigma_B$ 学习销售者 $s_j\in\Sigma_S$ 的历史所得的奖励折扣因子.

定义 **8**(协商时间向量). $Tm^{s_i}=\langle t_1^{s_i},t_2^{s_i},...,t_N^{s_i}\rangle$ 表示在 N 步之内选择 $s_i\in\Sigma_S$ 所对应的协商历史的协商时间向量.其中,$t_j^{s_i}$ 表示在第 j 步选择 s_i 所对应的协商历史的协商时间.

定义 **9**(信任向量). $T_S^{b_i}=\langle T_{b_i\to s_1},T_{b_i\to s_2},...,T_{b_i\to s_K}\rangle$ 表示 $b_i\in\Sigma_B$ 对各个 $s_j\in\Sigma_S$ 的一个信任向量.其中,$T_{b_i\to s_j}\in[0,1]$ 表示购买者 b_i 对销售者 s_j 的信任度.在信任依赖图中,如果$(b_i,s_j)\notin E_1$,则 $T_{b_i\to s_j}=R_{b_i\to s_j}$,其中,$T_{b_i\to s_j}$ 表示 s_j 相对于 b_i 的声誉度. $T_B^{b_i}=\langle r_{k_1}^{b_i},r_{k_2}^{b_i},...,r_{k_m}^{b_i}\rangle$ 表示 $b_i\in\Sigma_B$ 对其他 $b_j\in\Sigma_B$ 且 $i\neq j$ 的一个信任向量.其中,$m=|\Sigma_B|-1$ 且 $k_j\neq i$.

下面给出的算法 $EE(\alpha,\gamma)$是对 $E(\alpha,\gamma)$进行扩充,并将上述因素与奖励相结合起来.

算法. $EE(\alpha,\gamma)$.

在算法 $E(\alpha,\gamma)$的基础上将时间折扣因子和外部信息相结合. $\hat{ss}_j$ 表示折扣奖励和.

初始化:

对 $H(\alpha,n)$进行初始化, $\hat{ss}_j\leftarrow 0$.

For n=1,…, N do{

1. 执行 $H(\alpha,n)$获得分布向量 $Q(n)\in[0,1]^K$
2. 获得信任向量 $T_S^{b_i}\in[0,1]^K$,$P(n)\in[0,1]^K$ 是另一个分布向量,并且

   for j=1,…,K do {

   $p_j(n)\leftarrow q_j(n)\,T_{b_i\to s_j}$

3. 将所得分布与均匀分布叠加

$$\hat{p}_j(n)\leftarrow(1-\gamma)p_j(n)+\gamma/K.$$

$$\hat{r}^{s_j}(n)\leftarrow(\gamma/K)\times(r^{s_j}(n)/\hat{p}_j(n)).$$

4. 将第 3 步产生的奖励向量结合信任向量 $T_S^{b_i}\in[0,1]^K$ 有

$$\hat{r}^{s_j}(n)\leftarrow T_{b_i\to s_j}\times\hat{r}^{s_j}(n).$$

5. 结合时间折扣因子,并分别累计各个协商对象的奖励和

$$\hat{ss}_i\leftarrow\hat{ss}_i+\hat{p}_j(n)\times\hat{r}^{s_j}(n)\times(\delta_j)^{t-t_n^{s_j}}.$$

}

}

从上述算法可以看出:第 1 步从算法 $H(\alpha,n)$获得选择协商对象的分布 $Q(n)$;第 2 步结合信任向量产生新的分布;第 3 步将所得分布与均匀分布叠加;第 4 步将所得奖励与信任向量结合;第 5 步对各个协商对象所给的奖励进行时间折扣.算法结束后,选择最合适的协商对象:

$$S^*=\arg\max_{1\le i\le K}\left(\hat{ss}_i\right).$$

S^*表示与最大奖励和对应的协商对象.实验结果表明,这个结果要比前面的结果更精确,尽管考虑了上述因素,但算法复杂度依然是 $O(N^2)$.

4 实验与分析

本文安排的实验主要涉及 4 个购买者,即$|\Sigma_B|$=4;6 个销售者,即$|\Sigma_S|$=6.实验分如下几个步骤进行:(1) 只利用信任-声誉模型来选择协商对象;(2) 只利用最大奖励和来选择协商对象,分析α和γ;(3) 将信任-声誉模型与最大奖励和结合选择对象,分析α和γ.在实验中,为 6 个销售者分别准备了 100 个协商线程,共计 600 个协商线程.

4.1 实验1

这部分实验是以信任-声誉模型为工具,以购买者自己已有的协商历史为基础,建立信任依赖图.协商对象选择过程 Ch_P 执行结果的信任依赖图如图 1 所示.具体结果如下:

购买者之间的信任向量是:

$T_B^{b_1}=\langle 0.61,0.63,0.26\rangle$ 对应 $b_2b_3b_4$,

$T_B^{b_2}=\langle 0.47,0.31,0.22\rangle$ 对应 $b_1b_3b_4$,

$T_B^{b_3}=\langle 0.23,0.42,0.35\rangle$ 对应 $b_1b_2b_4$,

$T_B^{b_4}=\langle 0.20,0.19,0.61\rangle$ 对应 $b_1b_2b_3$.

购买者对销售者的信任向量是:

$T_S^{b_1}=\langle r_4^{b_1}t_2,r_4^{b_1}t_3,r_2^{b_1}t_4,t_1+r_3^{b_1}t_5,r_2^{b_1}t_6+r_3^{b_1}t_7,r_4^{b_1}t_8\rangle=\langle 0.16,0.09,0.44,0.75,0.84,0.20\rangle$,

$T_S^{b_2}=\langle r_4^{b_2}t_2,r_4^{b_2}t_3,t_4,r_1^{b_2}t_1+r_3^{b_2}t_5,t_6+r_3^{b_2}t_7,r_4^{b_2}t_8\rangle=\langle 0.13,0.07,0.87,0.36,1,0.17\rangle$,

$T_S^{b_3}=\langle r_4^{b_3}t_2,r_4^{b_3}t_3,r_2^{b_3}t_4,t_5+r_1^{b_3}t_1,t_7+r_2^{b_3}t_6,r_4^{b_3}t_8\rangle=\langle 0.21,0.12,0.37,0.51,0.84,0.27\rangle$,

$T_S^{b_4}=\langle t_2,t_3,r_2^{b_4}t_4,r_1^{b_4}t_1+r_3^{b_4}t_5,r_2^{b_4}t_6+r_3^{b_4}t_7,t_8\rangle=\langle 0.61,0.34,0.17,0.34,0.45,0.77\rangle$.

从上面 4 个信任向量可以看出,b_1,b_2,b_3 都选择 s_5 进行协商;b_4 选择 s_6 进行协商.从结果可以看出,b_2,b_3,b_4 都选择了和自己信念一致的结果,只有 b_1 由于声誉度而选择了 s_5.s_5 具有较高的选择概率.

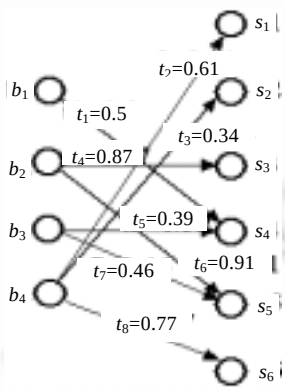

Fig.1 Trust dependent diagram for experiment

图 1 实验信任依赖图

4.2 实验2

这部分实验以 s_1,s_2,s_3,s_4,s_5 和 s_6 的协商历史为基础,通过模拟协商分别阅读它们的协商历史.实验在 100 个线程中动态找到 3 个稳定点:第 1 个稳定点维持在 N=6 步之内,在此期间,α=1.3903,γ=0.69517,并且 s_1 具有较高的选择概率;第 2 个稳定点维持在 N=28 步之内,在此期间,α=1.1035,γ=0.55176,并且 s_5 具有较高的选择概率,这个结果和实验 1 的结果相当;第 3 个稳定点维持在 N=100 步之内,它覆盖了所有的样本线程,在此期间,α=0.87586,γ=0.43793,并且 s_3 具有较高的选择概率,如图 2 所示.第 3 个稳定点持续的步骤数最多,支持它的历史信息比前两个稳定点更充分,结果更准确.但是,上面 3 个稳定点由于覆盖的步数和样本范围不同,所以所得的结果不同,为此,我们进行了第 3 步实验.

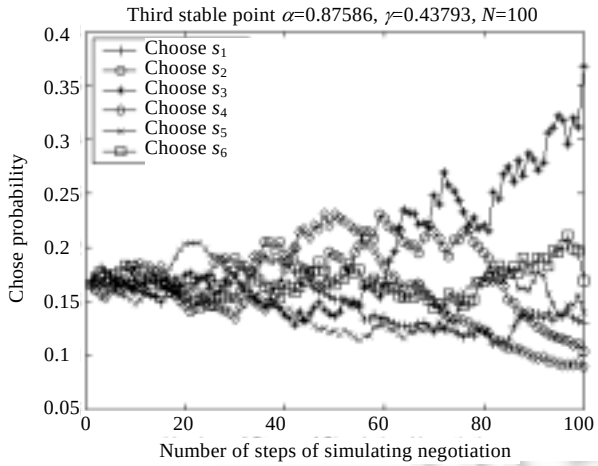

Fig.2 The third stable point in experiment 2

图 2 实验 2 中的第 3 个稳定点

4.3 实验3

这部分实验是考虑了第 3.5 节提到的几个因素而进行的实验.实验将信任向量 $T_S^{b_1}$, $T_S^{b_2}$, $T_S^{b_3}$, $T_S^{b_4}$ 和协商历史时间向量 Tm^{s_1}, Tm^{s_2}, Tm^{s_3}, Tm^{s_4}, Tm^{s_5}, Tm^{s_6} 以及时间折扣因子 δ^{b_1}, δ^{b_2}, δ^{b_3}, δ^{b_4} 结合到实验中.实验动态地找到 4 个稳定点:第 1 个稳定点维持在 N=5 步之内,且α=1.3903,γ=0.69517;第 2 个稳定点维持在 N=23 步之内,且α=1.1035,γ=0.55176;第 3 个稳定点维持在 N=93 步之内,且α=0.87586,γ=0.43793,如图 3 所示.第 4 个稳定点维持在 N=100 步之内,且α=0.69517,γ=0.34759,如图 4 所示.从这 4 个稳定点可以看出,它们都统一显示了 s_5 具有最高的选择概率.4 个稳定点所显示的次优结果都是 s_3.由于实验 3 的信息比实验 2 的信息充分,所以,实验 3 也正说明了 s_3 在实验 2 中是最优解.实验 2 和实验 3 的前 3 个稳定点的α,γ值是一致的,即它们是相同的稳定点,只是维持的步数稍有不同,这说明实验 3 更优于实验 2.另外,从实验 3 还可以看出,在相关信息辅助下,并在近期协商历史不太长的情况下,只要动态地找到第 1 个稳定点,就可以从指定的选择范围内选择出最好的协商对象.

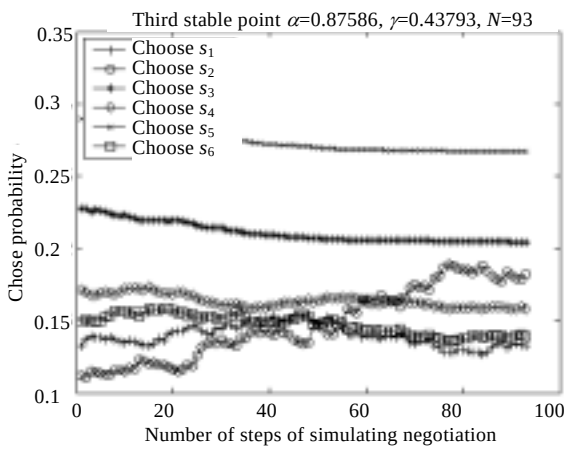

Fig.3 The third stable point in experiment 3

图 3 实验 3 中的第 3 个稳定点

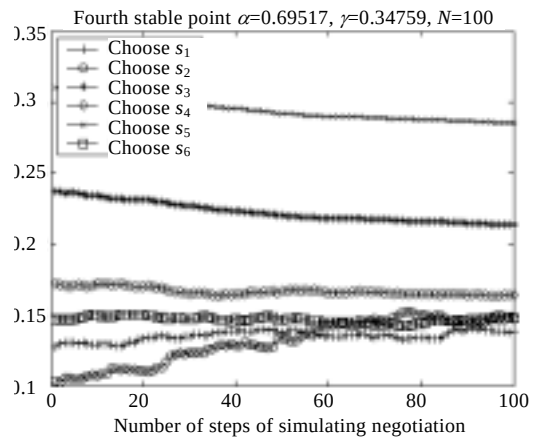

Fig.4 The fourth stable point in experiment 3

图 4 实验 3 中的第 4 个稳定点

5 结 论

多问题协商是一个复杂的动态交互过程.为了提高多问题协商过程的成功率,本文从购买者的观点重点解决协商前对销售者的选择问题.主要从以下几个方面来解决该问题:(1) 提出一个信任-声誉模型.在购买者和销售者的一个特定范围内,根据他们之间的协商历史,建立信任-声誉模型,购买者可以通过信任度和声誉度来选择销售者,但这种方法的缺陷是在特定范围内的协商历史信息不足.(2) 提出几种确定奖励分布的修改算法.以模拟协商为基础充分学习销售者已有的协商历史信息,结合 K 臂赌博机问题求解技术,将销售者选择问题转换为 K 臂赌博机问题来求解,并在无分布假设的情况下,在α和γ的约束下,动态产生选择各个销售者的概率分布.按照奖励最大化选择协商对象.(3) 在将(1)和(2)紧密结合的情况下,又结合了协商历史的时间因素,提出了相应的算法,提高了方法的实用性和有效性.实验分几个阶段进行,并证明了上述方法的有效性.在其他协商文献中很少涉及这个问题,本文较好地解决了协商前销售者的选择问题.

References:

[1] Coehoorn RM, Jennings NR. Learning an opponent's preferences to make effective multi-issue negotiation trade-offs. In: Henk GS, ed. Proc. of the 6th Int'l Conf. on e-Commerce. Delft: IEEE Press, 2004. 59−68.

[2] Wang LM, Huang HK, Chai YM. A learning-based multistage negotiation model. In: Yeung D, ed. Proc. of the 2004 Int'l Conf. on Machine Learning and Cybernetics. Shanghai: IEEE Press, 2004. 140−145.

[3] Faratin P, Sierra C, Jennings NR. Using similarity criteria to make issue trade-offs in automated negotiations. Artificial Intelligence, 2002,142(2):205−237.

[4] Guo Q, Chen C. An integrated-utility based optimization in multi-issue negotiation. Journal of Software, 2004,15(5):706−711 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/706.htm

[5] Wang LC, Chen SF. A multi-Agent multi-issue negotiation model. Journal of Software, 2002,13(8):1637−1643 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/1637.pdf

[6] Jøsang A. Trust-Based decision making for electronic transactions. In: Yngstrom L, Svensson T, eds. Proc. of the 4th Nordic Workshop on Secure Computer Systems. Kista: Stockholm University Press, 1999. 1−21.

[7] Dimitrakos T. Towards a formal model of trust in e-commerce. In: Matwin S, ed. Proc. of the AI-2001 Workshop on Novel E-Commerce Applications of Agents. Ottawa: Springer-Verlag, 2001. 1−10.

[8] Braynov S, Sandholm T. Contracting with uncertain level of trust. Computer Intelligence, 2002,18(4):501−514.

[9] Auer P, Cesa-Bianchi N, Fischer P. Finite-Time analysis of the multiarmed bandit problem. Machine Learning, 2002,47:235−256.

[10] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997,55(1):119−139.

[11] Auer Pr, Cesa-Bianchi N, Freund Y, Schapire RE. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In: Proc. of the 36th Annual Symp. on Foundations of Computer Science. New York: ACM Press, 1995. 322−331.

[12] Karl K, Iouri L. A model for multi-lateral negotiations on an Agent-based job marketplace. Electronic Commerce Research and Application, 2005,4(3):187−203.

[13] Dunne PE, Wooldridge M, Laurence M. The complexity of contract negotiation. Artificial Intelligence, 2005,164(1-2):23−46.

附中文参考文献:

[4] 郭庆,陈纯.基于整合效用的多议题协商优化.软件学报,2004,15(5):706−711. http://www.jos.org.cn/1000-9825/15/706.htm

[5] 王立春,陈世福.多 Agent 多问题协商模型.软件学报,2002,13(8):1637−1643. http://www.jos.org.cn/1000-9825/13/1637.pdf

王黎明(1963－),男,河南浚县人,博士,副教授,主要研究领域为分布式人工智能,数据挖掘,机器学习.

柴玉梅(1964－),女,副教授,主要研究领域为人工智能,机器学习.

黄厚宽(1940－),男,教授,博士生导师,CCF 高级会员,主要研究领域为分布式人工智能,机器学习,数据挖掘,演化计算.