文章编号:1672 - 058X(2011)05 - 0509 - 05

区间值模糊目标信息系统的规则提取

郭翠峰[1], 胡 鹏[2], 胡展闳[1]
(1. 西华大学 数学与计算机学院,成都 610039;2. 重庆师范大学 数学学院,重庆 401331)

摘 要:在区间值信息系统中定义了一种新的变精度相容关系,研究了这种相容关系诱导的极大变精度相容类;提出了基于极大变精度相容类的区间值模糊目标信息系统的粗糙集模型;进一步讨论了利用粗糙隶属函数获得决策规则的方法,最后通过实例说明了方法的可行性.

关键词:区间值信息系统;极大变精度相容类;粗糙集;区间值模糊目标信息系统;决策规则
中图分类号:TP18 **文献标志码**: A

粗糙集理论[1-2]是经典集合理论的推广,是近年发展起来的一种处理不确定性、不精确性和模糊性知识的有用数学工具. 经典粗糙集模型是基于完备信息系统上的等价关系,对无法用已知知识描述的对象集给出上、下近似. 经典粗糙集模型也可以推广到非等价关系(如相容关系、优势关系、偏序关系等)[3-12],这种推广可以用来处理不完备信息系统上的知识获取与推理. 但是,这些粗糙集方法主要处理的是离散属性值,而在很多实际问题中,由于问题的复杂性,往往会遇到条件属性值域是区间,甚至目标属性值域是模糊的情形. 条件属性值域是区间的信息系统称为区间值信息系统,许多学者已进行了深入的研究[6-10];目标属性为模糊的信息系统称为模糊目标信息系统,也有许多学者做过详细探讨[11-15]. 针对区间值模糊目标信息系统,Sun Bingzhen[11]等研究了区间值模糊目标信息系统的模糊粗糙集理论,Gong Zengtai[12]等提出了区间值模糊信息系统知识约简的方法. 然而,这些研究都是基于经典粗糙集模型下的等价关系. 鉴于此,在区间值信息系统中定义了一种新的变精度相容关系,讨论了这种变精度相容关系下的区间值模糊目标信息系统的规则获取方法.

1 区间值信息系统的基本概念和极大变精度相容类

定义 1[6] 设 $K=(U,A,V,F)$ 是一个信息系统,其中 $U=\{x_1,x_2,\cdots,x_n\}$ 为有限对象集,称为论域,每个 $x_i(i\leqslant n)$ 称为一个对象;$A=\{a_1,a_2,\cdots,a_m\}$ 为有限属性集,每个 $a_k(k\leqslant m)$ 称为一个属性;$F=\{f_k:k\leqslant m\}$ 为对象属性值映射集,其中 $f_k:U\rightarrow V_k(k\leqslant m)$,$V_k$ 是属性 a_k 的有限值域,$V=\bigcup_{a_k\in A}V_k$. 若 $\forall a_k\in A$, $\forall x_i\in U$,$f_k(x_i)$ 都是唯一确定的,则称信息系统 K 是完备的,否则便称它是不完备的;若 $f_k(x_i)=[l_i^k,r_i^k]=\{p\mid l_i^k\leqslant p\leqslant r_i^k\}\in I([0,1])$,则称 K 是区间值信息系统,其中 $I([0,1])$ 表示区间$[0,1]$上的闭区间全体.

定义 2 设 $K=(U,A,V,F)$ 是区间值信息系统,$a_k\in A,x_i,x_j\in U,f_k(x_i)=[l_i^k,r_i^k],f_k(x_j)=[l_j^k,r_j^k]$,称

$$S_{ij}^k=\begin{cases}1, & l_i^k=r_i^k=l_j^k=r_j^k\\ 1-\min\left\{\dfrac{|l_i^k-l_j^k|+|r_i^k-r_j^k|}{\max\{r_i^k,r_j^k\}-\min\{l_i^k,l_j^k\}},1\right\}, & \text{其他}\end{cases}$$

是对象 x_i,x_j 相对于属性 a_k 的相似度.

收稿日期:2011 - 03 - 11;修回日期:2011 - 04 - 18.
* 基金项目:西华大学无线电信号智能处理重点实验室基金(XZD0818-09).
作者简介:郭翠峰(1985 -),女,山东邹平人,硕士研究生,从事智能信息处理研究.

定义 3 设 $K=(U,A,V,F)$ 是区间值信息系统,$B\subseteq A$,$\alpha\in(0,1]$,称 $T_B^\alpha=\{(x_i,x_j)\in U\times U:\forall a_k\in B,S_{ij}^k\geqslant\alpha\}$ 是区间值信息系统上的变精度相容关系,阈值 α 称为相似水平.

记 $T_B^\alpha(x_i)=\{x_j\in U:(x_i,x_j)\in T_B^\alpha\}$,则 $(x_i,x_j)\in T_B^\alpha$ 表示对象 x_i 与 x_j 对于 B 中任一个属性的相似程度都不低于给定的阈值 α.

定理 1 设 $K=(U,A,V,F)$ 是区间值信息系统, T_B^α 是如上定义的二元关系,则有:

(1) T_B^α 是 U 上的相容关系,即 T_B^α 是自反、对称的;

(2) 若 $B_1\subseteq B_2\subseteq A$,则对 $\alpha\in(0,1]$,有 $T_A^\alpha\subseteq T_{B_2}^\alpha\subseteq T_{B_1}^\alpha$,$T_A^\alpha(x_i)\subseteq T_{B_2}^\alpha(x_i)\subseteq T_{B_1}^\alpha(x_i)$;

(3) 对任意 $B\subseteq A$,$T_B^\alpha=\bigcap\limits_{a_k\in B}T_{\{a_k\}}^\alpha$;

(4) 若 $0<\alpha\leqslant\beta\leqslant1$,则有 $T_B^\beta\subseteq T_B^\alpha$,$T_B^\beta(x_i)\subseteq T_B^\alpha(x_i)$.

证明 (1)(2)(3)由 T_B^α 的定义进行验证即可,下面证明(4).

对任意 $(x_i,x_j)\in T_B^\beta$,由定义 3,$\forall a_k\in B$ 就有 $S_{ij}^k\geqslant\beta$,而 $\alpha\leqslant\beta$,所以 $\forall a_k\in B$ 有 $S_{ij}^k\geqslant\alpha$,于是 $(x_i,x_j)\in T_B^\alpha$,从而 $T_B^\beta\subseteq T_B^\alpha$ 成立. 再由 T_B^α 与 $T_B^\alpha(x_i)$ 的关系即得 $T_B^\beta(x_i)\subseteq T_B^\alpha(x_i)$.

区间值信息系统 $K=(U,A,V,F)$ 上的变精度相容关系 T_B^α 是自反的,因而对 $B\subseteq A$,集合 $\{T_B^\alpha(x_i);x_i\in U\}$ 构成论域 U 的一个覆盖. 但变精度相容关系 T_B^α 一般不是等价关系,利用 $\{T_B^\alpha(x_i);x_i\in U\}$ 作为基本知识,有以下不足[10]:一是 $T_B^\alpha(x_i)$ 中所有对象虽然都与 x_i 在相似水平 α 下关于属性集 B 相容,但未必两两之间在相似水平 α 下关于 B 相容;二是可能出现 $T_B^\alpha(x_i)\subseteq T_B^\alpha(x_j)$ 的情形. 因此,为了解决以上两问题,基于 Leung[10] 定义的极大一致块,给出区间值信息系统上极大变精度相容类的定义.

定义 4 设 $K=(U,A,V,F)$ 是区间值信息系统, α 是相似水平,$B\subseteq A$,若 $M\subseteq U$ 是满足其中任何两个对象关于属性集 B 在相似水平 α 下都相容,则称 M 是一个 T_B^α 变精度相容类;若 M 是一个 T_B^α 变精度相容类,且 $\forall x_i\in U-M$,$\exists x_j\in M$ 使 $(x_i,x_j)\notin T_B^\alpha$,则称 M 是一个 T_B^α 极大变精度相容类. 以 $K_B^\alpha(U)$ 记 U 上所有 T_B^α 极大变精度相容类的集合,则易知 $K_B^\alpha(U)$ 是 U 的一个覆盖.

2 粗糙集模型建立

定义 5 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,其中 (U,A,V,F) 是区间值信息系统,$D=\{d_1,d_2,\cdots,d_p\}$ 为非空有限目标属性集,每个 $d_l(l\leqslant p)$ 称为一个目标属性;$G=\{g_l:l\leqslant p\}$ 为模糊目标属性值映射集,其中 $g_l:U\rightarrow W_l$,$W_l\in F(U)$,$W=\bigcup\limits_{1\leqslant l\leqslant p}W_l$,$W_l(x_i)\in[0,1]$,$i\leqslant n$,$l\leqslant p$, $F(U)$ 为 U 上模糊集合的全体.

定义 6 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,$B\subseteq A$,T_B^α 是区间值信息系统上的变精度相容关系,$\forall x_i\in U$,包含 x_i 的相容类为 $T_B^\alpha(x_i)=\{x_j\in U:(x_i,x_j)\in T_B^\alpha\}$,对任意的模糊集 $W\in F(U)$,令 $\overline{T_B^\alpha}(W)(x_i)=\max\{W(x_j):x_j\in T_B^\alpha(x_i)\}$,$\underline{T_B^\alpha}(W)(x_i)=\min\{W(x_j):x_j\in T_B^\alpha(x_i)\}$,其中 $W(x_j)$ 表示对象 x_j 的隶属函数值. 则 $\overline{T_B^\alpha}(W)$,$\underline{T_B^\alpha}(W)$ 分别称为模糊集 W 在区间值模糊目标信息系统 K 下基于变精度相容关系 T_B^α 的模糊上近似集与模糊下近似集,其中 $\overline{T_B^\alpha}:F(U)\rightarrow F(U)$ 与 $\underline{T_B^\alpha}:F(U)\rightarrow F(U)$ 分别称为区间值模糊目标信息系统的模糊上近似算子与模糊下近似算子. 显然,$\overline{T_B^\alpha}(W)$ 与 $\underline{T_B^\alpha}(W)$ 是一对模糊集合.

定理 2 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,$W\in F(U)$,$B\subseteq A$,则有:

(1) $\underline{T_B^\alpha}(W)\subseteq W\subseteq\overline{T_B^\alpha}(W)$;(2) 若 $W_1\subseteq W_2$,则 $\underline{T_B^\alpha}(W_1)\subseteq\underline{T_B^\alpha}(W_2)$,$\overline{T_B^\alpha}(W_1)\subseteq\overline{T_B^\alpha}(W_2)$.

证明 由定义 6 知显然成立.

定义 7 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,$W\in F(U)$,在相容关系 T_A^α 下定义其相容粗糙集为 $T_A^\alpha(W)=(\underline{T_A^\alpha}(W),\overline{T_A^\alpha}(W))$,对 $\forall x_i\in U$,极大变精度相容类 $M\in K_B^\alpha(U)$ 在相容粗糙集中的隶属函数为 $\mu_{\underline{T_A^\alpha}(W)}(M_i)=\min\{W(x_i):x_i\in M_i\}$,$\mu_{\overline{T_A^\alpha}(W)}(M_i)=\max\{W(x_i):x_i\in M_i\}$.

定义 8 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,$W\in F(U)$,对象 $x_i\in U$ 对 W 的相容粗糙模糊隶属函数定义为 $\eta_W(x_i)=\dfrac{|M\cap W|}{|M|}$,其中 $M\in K_A^\alpha(U)$ 为 x_i 的极大变精度相容类.

定理 3 设 $K=(U,A,V,F;D,W,G)$ 为区间值模糊目标信息系统,$W\in F(U)$,$\eta_W(x_i)$ 是如上定义的相容粗糙模糊隶属函数,则有:

(1) $0 \leqslant \eta_W(x_i) \leqslant 1$；(2) $\forall A,B \subseteq U$，$\eta_{A\cup B}(x_i) \geqslant \max\{\eta_A(x_i),\eta_B(x_i)\}$；(3) $\forall A,B \subseteq U$，$\eta_{A\cap B}(x_i) \leqslant \min\{\eta_A(x_i),\eta_B(x_i)\}$.

证明　(1) 因为 $\varnothing \leqslant M\cap W \leqslant M$，所以 $0 \leqslant \eta_W(x_i) \leqslant 1$ 成立；

(2) $\eta_{A\cup B}(x_i) = \dfrac{|M\cup(A\cup B)|}{|M|} = \dfrac{|(M\cup A)\cup(M\cup B)|}{|M|} \geqslant \dfrac{|M\cup A|}{|M|} = \eta_A(x_i)$，同理可证 $\eta_{A\cup B}(x_i) \geqslant \eta_B(x_i)$；

(3) $\eta_{A\cap B}(x_i) = \dfrac{|M\cup(A\cap B)|}{|M|} = \dfrac{|(M\cup A)\cap(M\cup B)|}{|M|} \leqslant \dfrac{|M\cup A|}{|M|} = \eta_A(x_i)$，同理可证 $\eta_{A\cap B}(x_i) \leqslant \eta_B(x_i)$.

3　决策规则的获取

定义 9　设 $K=(U,A,V,F)$ 是区间值信息系统，α 是相似水平，$B\subseteq A$，对 $M\in K_B^\alpha(U)$，记 $\mathrm{des}\, M = \bigwedge_{a_k\in B}(a_k, [\max\{l_i^k | x_i\in M\}, \min\{r_i^k | x_i \in M\}])$，称 des M 是区间值信息系统的极大变精度相容类 M 的属性特征描述.

利用相容粗糙隶属函数可以得到区间值模糊目标信息系统中每一个对象在不同决策类中的可能值，也就能获得所需的决策规则，类似于文献[13]具体算法简述如下：

输入：区间值模糊目标信息系统；

输出：决策规则；

(1) 求出论域 U 上所有极大变精度相容类的集合 $K_A^\alpha(U)=(M_1,M_2,\cdots,M_n)$，利用定义 9 求出每个极大变精度相容类的属性特征描述 des $M_i(1\leqslant i\leqslant n)$；

(2) 求下近似对各决策属性 d_l 的粗糙模糊隶属函数：$\mu_{\overline{T_A^\alpha}(d_l)}(K_A^\alpha(U)) = (\mu_{\overline{T_A^\alpha}(d_l)}(M_1), \mu_{\overline{T_A^\alpha}(d_l)}(M_2),\cdots,\mu_{\overline{T_A^\alpha}(d_l)}(M_n))$；

(3) 求上近似对各决策属性 d_l 的粗糙模糊隶属函数：$\mu_{\overline{T_A^\alpha}(d_l)}(K_A^\alpha(U)) = (\mu_{\overline{T_A^\alpha}(d_l)}(M_1), \mu_{\overline{T_A^\alpha}(d_l)}(M_2),\cdots,\mu_{\overline{T_A^\alpha}(d_l)}(M_n))$；

(4) 利用 $\eta_{d_l}(M_i) = \dfrac{|M_i \cap d_l|}{|M_i|} = \dfrac{\left|\sum\limits_{x_i\in U}\min(\mu_{M_i}(x_i), W_{d_l}(x_i))\right|}{|M_i|}$ 求决策类的相容粗糙隶属函数值：$\eta_{d_l}(K_A^\alpha(U)) = (\eta_{d_l}(M_1),\eta_{d_l}(M_2),\cdots,\eta_{d_l}(M_n))$，其中 $\mu_{M_i}(x_i) = \begin{cases}1, & x_i \in M_i\\ 0, & x_i \notin M_i\end{cases}$；

(5) 规则提取：若对象 x_i 在条件属性集 $A=\{a_1,a_2,\cdots,a_m\}$ 上的取值 $(a_1(x_i),a_2(x_i),\cdots,a_k(x_i)) = \mathrm{des}M_i$，则它属于决策类 d_l 的可能性为 $\eta_{d_l}(M_i)$.

例　考虑如下区间值模糊目标信息系统，其中论域 $U=\{x_1,x_2,\cdots,x_6\}$，条件属性集 $A=\{a_1,a_2,a_3,a_4\}$，目标属性集 $D=\{d_1,d_2\}$，条件属性值 $a_k(x_i)=[l_i^k,r_i^k]$ 是区间值，如 $a_1(x_2)=[0.8,1.0]$，决策属性值 $d_l(x_i)\in[0,1]$，如 $d_1(x_2)=0.9$.

表 1　区间值模糊目标信息系统

	a_1	a_2	a_3	a_4	d_1	d_2
x_1	[0.5,0.8]	[0.2,0.5]	[0.1,0.5]	[0.4,0.7]	0.4	0.7
x_2	[0.8,1.0]	[0.4,0.6]	[0.3,0.6]	[0.4,0.6]	0.9	0.2
x_3	[0.6,0.8]	[0.3,0.6]	[0.1,0.4]	[0.3,0.8]	0.5	0.8
x_4	[0.6,0.9]	[0.3,0.5]	[0.2,0.5]	[0.4,0.5]	0.1	0.9
x_5	[0.7,1.0]	[0.4,0.5]	[0.2,0.6]	[0.4,0.6]	0.6	0.3
x_6	[0.8,0.9]	[0.5,0.6]	[0.3,0.7]	[0.4,0.7]	0.7	0.4

(1) 取相似水平 $\alpha=0.5$，则根据定义 2、3 及 4 可以算出论域 U 上的极大变精度相容类为 $K_A^{0.5}(U)=\{M_1,M_2,M_3,M_4\}$，其中 $M_1=\{x_1,x_3\}$，$M_2=\{x_2,x_5\}$，$M_3=\{x_2,x_6\}$，$M_4=\{x_4,x_5\}$.

(2) 极大变精度相容类的属性特征描述如下：

des $M_1=(a_1,[0.6,0.8])\wedge(a_2,[0.3,0.5])\wedge(a_3,[0.1,0.4])\wedge(a_4,[0.4,0.7])$

des $M_2=(a_1,[0.8,1.0])\wedge(a_2,[0.4,0.5])\wedge(a_3,[0.3,0.6])\wedge(a_4,[0.4,0.6])$

des $M_3=(a_1,[0.8,0.9])\wedge(a_2,[0.5,0.6])\wedge(a_3,[0.3,0.6])\wedge(a_4,[0.4,0.6])$

des $M_4=(a_1,[0.7,0.9])\wedge(a_2,[0.4,0.5])\wedge(a_3,[0.2,0.5])\wedge(a_4,[0.4,0.5])$

(3) 求下近似的隶属函数:对于目标属性 d_1,有:$\mu_{\underline{T_A^\alpha}(d_1)}(M_1)=\min\{W_{d_1}(\{x_1,x_3\})\}=\min\{0.4,0.5\}=0.4$,同理可得 $\mu_{\underline{T_A^\alpha}(d_1)}(M_2)=0.6$,$\mu_{\underline{T_A^\alpha}(d_1)}(M_3)=0.7$,$\mu_{\underline{T_A^\alpha}(d_1)}(M_4)=0.1$,从而 $\mu_{\underline{T_A^\alpha}(d_1)}(K_A^\alpha(U))=(0.4,0.6,0.7,0.1)$;同理 $\mu_{\underline{T_A^\alpha}(d_2)}(K_A^\alpha(U))=(0.7,0.2,0.2,0.3)$.

(4) 求上近似隶属函数:与下近似隶属函数的求法类似,可得:

$$\mu_{\overline{T_A^\alpha}(d_1)}(K_A^\alpha(U))=(0.5,0.9,0.9,0.6),\mu_{\overline{T_A^\alpha}(d_2)}(K_A^\alpha(U))=(0.8,0.3,0.4,0.9)$$

(5) 利用 $\eta_{d_l}(M_i)=\dfrac{|M_i\cap d_l|}{|M_i|}=\dfrac{|\sum_{x_i\in U}\min(\mu_{M_i}(x_i),W_{d_l}(x_i))|}{|M_i|}$ 可求得决策类的相容粗糙模糊隶属函数值为:$\eta_{d_1}(M_1)=0.45$,$\eta_{d_2}(M_1)=0.75$;$\eta_{d_1}(M_2)=0.75$,$\eta_{d_2}(M_2)=0.25$,$\eta_{d_1}(M_3)=0.8$,$\eta_{d_2}(M_3)=0.3$;$\eta_{d_1}(M_4)=0.35$,$\eta_{d_2}(M_4)=0.6$.

(6) 规则提取:若 $(a_1(x),a_2(x),a_3(x),a_4(x))=([0.6,0.8],[0.3,0.5],[0.1,0.4],[0.4,0.7])$,则 x 支持决策属性 d_1 的可能性为 0.45,x 支持决策属性 d_2 的可能性为 0.75;

若 $(a_1(x),a_2(x),a_3(x),a_4(x))=([0.8,1.0],[0.4,0.5],[0.3,0.6],[0.4,0.6])$,则 x 支持决策属性 d_1 的可能性为 0.75,x 支持决策属性 d_2 的可能性为 0.25;

若 $(a_1(x),a_2(x),a_3(x),a_4(x))=([0.8,0.9],[0.5,0.6],[0.3,0.6],[0.4,0.6])$,则 x 支持决策属性 d_1 的可能性为 0.8,x 支持决策属性 d_2 的可能性为 0.3;

若 $(a_1(x),a_2(x),a_3(x),a_4(x))=([0.7,0.9],[0.4,0.5],[0.2,0.5],[0.4,0.5])$,则 x 支持决策属性 d_1 的可能性为 0.35,x 支持决策属性 d_2 的可能性为 0.6.

参考文献:

[1] PAWLAK Z. Rough set[J]. International of Computer and Information Science,1982(11):341-356

[2] PAWLAK Z. Rough set:Theoretical aspects of reasoning about data[M]. Boston:Kluwer Academic Publishers,1991

[3] 张文修,梁怡,吴伟志. 信息系统与知识发现[M]. 北京:科学出版社,2003

[4] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社,2001

[6] 陈子春,秦克云. 区间值信息系统在变精度相容关系下的属性约简[J]. 计算机科学,2009,36(3):163-166

[7] 陈子春,秦克云. 区间值信息系统基于极大相容类的属性约简[J]. 模糊系统与数学,2009,23(6):126-132

[8] QIQN Y H, LIANG J. Dang Chuangyin. Interval ordered information systems[J]. Computer and Mathematics with Applications, 2008,56(8):1994-2009

[9] 张楠,苗夺谦,岳晓冬. 区间值信息系统的知识约简[J]. 计算机研究与发展,2010,47(8):1362-1371

[10] LEUNG Y, FISCHER M, WU W Zi, et al. A rough set approach for the discovery of classification rules in interval-valued information systems[J]. International Journal of Approximate Reasoning,2008,47(2):233-246

[11] SUN B ZH, GONG Z T, CHEN DEGANG. Fuzzy rough set theory for the interval-valued fuzzy information systems[J]. Information Sciences,2008,178:2794-2815

[12] GONG Z Ti, SUN B ZH, CHEN DEGANG. Rough set theory for the interval-valued fuzzy information systems[J]. Information Sciences,2008,178:1968-1985

[13] 梁小芝,王伟平. 基于最大相容类的不完备模糊目标信息系统规则提取[J]. 湖南科技学院学报,2008,29(4):110-114

[14] 魏大宽,黄兵,周献中. 不完备模糊目标信息系统粗集模型与知识约简[J]. 计算机工程,2006,32(8):48-51

[15] 魏大宽,周献中,黄兵. 不完备模糊决策信息系统的粗集模型与精度约简[J]. 计算机科学,2006,33(6):182-185

[16] 徐裕生,杜英阈,王佳佳. 求解有约束非线性规划问题的新算法[J]. 重庆理工大学学报,2010,24(6):101-103

Rule Acquisition in Interval-valued Fuzzy Objective Information Systems

GUO Cui-feng[1] ,**HU Peng**[2] ,**HU Zhan-hong**[1]

(1. School of Mathematics and Computer, Xihua University, Chengdu 610039;
2. School of Mathematics, Chongqing Normal University, Chongqing 401331, China)

Abstract: A new variable precision tolerance relation is defined and the induced maximal variable precision tolerance classes in interval valued information system are studied, this paper proposes the rough set model based on the maximal variable precision tolerance classes in interval-valued fuzzy objective information systems. Furthermore, the approaches to decision rules acquisition based on membership function are discussed and finally an example illustrates that the approaches are valid.

Key words: interval-valued information systems; maximal variable precision tolerance classes; rough set; interval-valued fuzzy objective information systems; decision rules

责任编辑:代小红

(上接第 495 页)

Information Leaking and Information Discovery Advantage of Institutional Investors ——Evidence from Asset Injection of Listed Companies of China

LEI Qian-hua[1] , **LIU Jian-hua**[2] , **JI Hua**[3]

(1. School of Management, Sun Yat-sen University, Guangzhou 510275, China;
2. School of Lingnan, Sun Yat-sen University, Guangzhou 510275, China;
3. School of Accounting, Zhongnan University of Finance, Economics and Law, Wuhan 430073, China)

Abstract: Institutional investors mainly have two kinds of information advantage, one is private information discovery advantage in advance and the other is interpretation advantage of public information. The existed literature mostly concerns about the interpretation advantage of public information of institutional investors. Through separating transaction records of institutional investors from high-frequency transaction database for per account and by using a special group of samples of asset injection, this paper discusses private information discovery advantage of institutional investors in advance and reveals that information leaking is chiefly caused by private information of institutional investors and that institutional investors extraordinarily purchase before asset injection disclosure, which lead to the change of stock price. This paper provides early warning for investment behaviors of individual investors and provides reference for the management of China Securities Regulatory Commission.

Key words: institutional investor; information leaking; asset injection

责任编辑:代小红