经典的知识依赖性及属性重要性度量的新注记

陈 飞 姜 麟 李金海
(昆明理工大学理学院 昆明 650500)

摘 要 知识依赖性及属性重要性度量是粗糙集的重要概念,广泛应用于知识约简和规则提取等方面。经典的知识依赖性及属性重要性度量在处理数据方面有局限性,有时无法得到较为精确、合理的度量结果,从而导致后续应用中得到的结果出现一系列的偏差。因此,通过深度分析经典知识依赖性,结合多数包含关系,并加入可信系数,提出了一种新的知识依赖性及属性重要性度量方法。最后,将新度量方法应用于一个决策信息系统,分析结果表明新度量方法是有效的。

关键词 粗糙集,决策信息系统,知识依赖性,属性重要性,度量
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.057

New Note on Classical Measure of Knowledge Dependency and Attribute Significance

CHEN Fei JIANG Lin LI Jin-hai
(Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China)

Abstract Measure of knowledge dependency and attribute significance is an important issue in rough set theory. It has been widely applied to knowledge reduction, rule extraction, etc. Classical measure of knowledge dependency and attribute significance has a littie bit of limitation in dealing with data, and sometimes it cannot obtain precise and reasonable results, which leads to a series of deviations in the subsequent applications. To this end, through deep analysis of the existing classical knowledge dependency, combining it with the majority inclusion relation, and adding credibility parameters, a new method of measuring knowledge dependency and attribute significance was proposed. Finally, the new measure method was applied to decision information system. And the analysis results show that the new method is effective.

Keywords Rough set, Decision information system, Knowledge dependency, Attribute significance, Measure

1 引言

粗糙集是由波兰学者 Z. Pawlak 于 20 世纪 80 年代初提出的一种处理含糊、不精确、不确定知识的实用理论,其最主要的目的之一是在大数据中寻找有实际意义的决策规则来获取新知识。

在粗糙集理论中,把感兴趣的对象组成的有限集合 $U\neq\varnothing$ 称为论域[1],通过等价关系对 U 进行划分,在此基础上利用划分得到的等价类来分析不同的问题。比如在保持知识分类能力不变的情况下对信息系统进行属性约简[1];通过属性重要性度量来识别决策表中属性的重要性;通过正区域提取确定性规则,借助包含度刻画规则的精度等等。

在现实生活中,很多的信息系统在进行不可分辨关系的划分时,由于其具有不确定性,而一些经典粗糙集方法的处理能力有限,因此有关学者对一些经典粗糙集方法提出了各自的看法[2-4]。胡丹等[2]提出经典知识依赖性度量存在一定的局限性,有时得不到客观、合理的结果;连钢等[3]也提出在比较属性之间的重要性时往往会得到一些与实际不相吻合的结论;赵乃刚等[4]也认为在现实生活中条件属性值与决策属性值往往具有强烈的相关性,这种相关性通常表现为领域知识。为了改善经典知识依赖性度量在实际应用中的效果,张文修等[5]首先提出包含度理论,将"包含关系"度量化,解决了"关系"的不确定性,同时也给其它有关不确定性工作的改善做了铺垫。此外,为了深入理解粗糙集数据分析中的度量关系,梁吉业等[6]以包含度理论为基础,建立了包含度与粗糙集中几个经典度量之间的关系,证实了粗糙集数据分析中的有关度量都可以归结为包含度。然而,文献[6]只是说明了包含度与经典度量之间存在某种关系,并未从实质上对经典度量方法进行改善。于是,苗夺谦等[7-9]将信息论引入粗糙集理论中,进行了知识的信息表示与信息度量之间的研究,并讨论了知识的不确定性与信息熵之间的关系。但文献[8,9]提出的利用信息熵来处理不确定性的方法在某些情况下仍然无法区分属性重要性。另外,胡丹等[2]给出了知识依赖性度量的公理化定义,同时从依赖度的角度对知识的相对重要性进行了描述。

上述文献[2-9]均未提到当决策分类分别关于两个不同的条件分类的正域中元素个数相同时,哪个条件属性子集与决策属性依赖性更大的问题。针对此问题,本文在经典知识

到稿日期:2014-12-11 返修日期:2015-04-04 本文受国家自然科学基金(61305057),云南省教育厅基金(2010Y389)资助。
陈 飞(1990-),男,硕士生,主要研究领域为粗糙集、并行数据挖掘,E-mail:chenfei19900123@163.com;姜 麟(1966-),男,副教授,硕士生导师,主要研究领域为并行计算和智能计算;李金海(1984-),男,博士,副教授,主要研究领域为粗糙集、概念格与粒计算。

依赖性度量的基础上，使经典知识依赖性度量方法得到顺延的同时，提出一种更加合理的知识依赖性及属性重要性度量方法，即在经典知识依赖性度量公式中引入多数包含关系得到新的知识依赖性度量公式，再加入可信系数得到新的属性重要性度量公式，使得分析结果更加合理且可靠。此外，将改进后的方法应用于一个决策信息系统，其比较分析结果表明新度量方法是有效的。

2 Rough 集的相关理论

信息系统 $S=(U,A,V,f)$，$U=\{x_1,x_2,\cdots,x_n\}$是对象的非空有限集合，称为论域；A 是属性的非空有限集合；特别地，如果 $A=C\cup D$，C 为条件属性的非空有限集，D 为决策属性的非空有限集，且 $C\cap D=\emptyset$，则称 S 为决策信息系统。其中 $V=\bigcup_{a\in A}V_a$，V_a 是属性 a 的值域；$f:U\times A\to V$ 是一个信息函数，它为每一个对象的每个属性赋予一个信息值，即 $\forall a\in A$，$x\in U$，则有 $f(x,a)\in V_a$。若 $D=\emptyset$，则称信息系统为数据表，否则称为决策表。

论域 $U\neq\emptyset$ 中的任何概念族称为关于 U 的抽象知识，简称知识。U 上的一族划分称为关于 U 的一个知识库。

定义 1[1]（分类定义） 设 $S=(U,A,V,f)$是一个信息系统，$\forall a\in A$，可以引入 U 的一个划分 U/a：对于两个对象 u，$v\in U$，若 $a(u)=a(v)$，则 u，v 分在同一类。

定义 2[1,10]（条件分类和决策分类） 根据条件属性 C 对 U 的划分称为条件分类，可表示为 $U/C=\{X_1,X_2,\cdots,X_m\}$，其中每个元素为一个条件类。根据决策属性 D 对 U 的划分称为决策分类，可表示为 $U/D=\{Y_1,Y_2,\cdots,Y_n\}$，其中每个元素为一个决策类。

定义 3[1]（依赖度） 设 $K=(U,R)$是一个知识库，$P,Q\subseteq R$。如果知识 Q 依赖于知识 P，则在知识库中，知识 Q 是多余的，在这种情况下，知识 $P\cup Q$ 与知识 P 描述相同的对象特点。

定义 4[1]（依赖性度量公式） 知识的依赖性可以是部分的，则知识 Q 仅有部分是由知识 P 导出，部分可导出可通过正域来定义。令 $K=(U,R)$，且 $P,Q\subseteq R$，记

$$k=\gamma_P(Q)=|pos_P(Q)|/|U|$$

其中 $pos_P(Q)=\bigcup_{X\in U/Q}\underline{P}X$。

称知识 Q 是 $k(0\leqslant k\leqslant 1)$度依赖于知识 P，记作 $P\Rightarrow Q$（当 $k=1$ 时，称 Q 完全依赖于 P；当 $0<k<1$ 时，称 Q 部分依赖于 P；当 $k=0$ 时，称 Q 完全独立于 P）。

定义 5[1]（属性重要性） 设 $S=(U,A,V,f)$是一个信息系统，其中 $A=U\cup D$，假设 $\emptyset\subset P\subseteq C$，$\emptyset\subset Q\subseteq D$，$U/Q\neq\{U\}$。给定 $p\in P$，如果 $S_P(Q)\Leftrightarrow S_{P-\{p\}}(Q)$，则称 p 在 P 中关于 Q 是重要的；如果 $S_P(Q)=S_{P-\{p\}}(Q)$，则称 p 在 P 中关于 Q 是不重要的。

定义 6[1]（支持子集） 根据定义 5，可以引入 U 的一个划分 U/a，其中 $W\subseteq U$，定义 W 的下近似为 $W^{(U/p)^-}=\bigcup_{V\in U/a,V\subseteq W}V$，可用 $S_a(W)$来表示，也可用 $pos_a(W)$来表示。

定义 7[1]（重要性度量公式） 根据定义 5 及定义 6，p 在 P 中关于 Q 的重要性为

$$sig^Q_{P-\{p\}}(p)=(|S_P(Q)|-|S_{P-\{p\}}(Q)|)/|U|$$

3 一种新的知识依赖性及属性重要性度量的改进方案

3.1 经典的知识依赖性度量的改进

经典的知识依赖性度量在某些情况下不能对客观事实进行如实的反映[2]，只适合用于同一约简中各个属性之间的比较[3]，从而不能较好地刻画其属性重要性，这对后续工作（如属性约简、决策规则提取等）成果的可靠性会有影响。为此，本文提出了一种基于经典知识依赖性度量、以多数包含关系为前提的新的知识依赖性度量方法来使得分析结果更加有效。

粗糙集中用 k 来表示知识 Q 依赖于知识 P 的度量[1]，且依赖性可以作为进行聚类和约简算法的依据[11]。然而该依赖度存在不合理性，如文献[2]中例 1 提到的。

根据定义 2，有如下划分：

$U/\{c_2,c_3\}=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6,x_7,x_8,x_9\}\}$

$U/D=\{\{x_2,x_3,x_4,x_9\},\{x_1,x_5,x_6,x_7,x_8\}\}$

由定义 4 可知，$k=0$（即决策属性 D 完全独立于条件属性集 $P=\{c_2,c_3\}$），但是属性集 P 和属性 D 在对象 x_2-x_8 的划分上是相同的，因此，在直观上可以认为属性集 P 和属性 D 之间有很强的依赖性。而出现这种情况的主要原因是经典知识依赖性度量在实例应用中存在局限性。

针对上述问题，本文在经典知识依赖性度量的基础上引入了包含度的概念。

定义 8[6]（包含度） 设 U 是有限非空集合，$P(U)$表示 U 的所有子集构成的幂集，对任意 $A,B\in P(U)$，记

$$D_0(B/A)=\begin{cases}\dfrac{|A\cap B|}{|A|}, & A\neq\emptyset\\ 1, & A=\emptyset\end{cases}$$

则称 $D_0(B/A)$为 A 关于 B 的包含度，即 B 包含 A 的程度。这里 $|\cdot|$表示集合的元素个数。

定义 9[1]（相对错误分类率） 设 X 和 Y 表示有限论域 U 的非空子集。如果每一个 $e\in X$ 同时有 $e\in Y$，则称 Y 包含 X，记作 $Y\supseteq X$。令

$$c(X,Y)=\begin{cases}1-D_0(Y/X), & |X|>0\\ 0, & |X|=0\end{cases}$$

其中 $D_0(Y/X)=\dfrac{|X\cap Y|}{|X|}$，且 $|\cdot|$表示集合的元素个数。

定义 10[1]（多数包含关系） 令 $0\leqslant\beta<0.5$，有：

$$Y\overset{\beta}{\supseteq}X\Leftrightarrow c(X,Y)\leqslant\beta$$

成立，则称 X 与 Y 满足多数包含关系。

文献[1]在变精度粗糙集模型中提出一种属性近似依赖度量，即

$$r(P,Q,\beta)=|pos(P,Q,\beta)|/|U|$$

其中 $pos(P,Q,\beta)=\bigcup_{Y\in U/Q}\underline{ind(P)_\beta}Y$。

根据表 1 及其划分，可令 $U/\{c_2,c_3\}=\{X_1,X_2\}$，$U/D=\{Y_1,Y_2\}$，那么在多数包含关系的前提下（取 $\beta=0.5$），由定义 4 可求得 $k=1$（即决策属性 D 完全依赖于条件属性集 $P=\{c_2,c_3\}$）。但是在求 Y_1 关于 P 的依赖度时，对象 x_1 并不属于 Y_1 关于 P 的正域，且在实际应用中这种偏差会更大。因

此，这种属性的近似依赖性度量在实际应用中也存在不足。

针对以上几点，本文以多数包含关系为前提，提出一种新的知识依赖性度量方法。

定义 11 设 $K=(U,R)$ 是一个知识库，且有知识 $P=\{X_1,X_2,\cdots,X_n\}$，知识 $Q=\{Y_1,Y_2,\cdots,Y_n\}$，如果 P 和 Q 满足多数包含关系（即满足 $Q\overset{0.5}{\supseteq}P$），则有

$$k'=\gamma_P{}'(Q)=|pos_P{}'(Q)|/|U|=|S_P{}'(Q)|/|U|$$

其中，$|pos_P{}'(Q)|=\sum_{i\in|P|,j\in|Q|}[1-c(X_i,Y_j)]|X_i|$。

最后，再根据表 1 及定义 11 可求得 $k'=7/9$（即决策属性 D 与条件属性集 $P=\{c_2,c_3\}$ 之间的依赖度为 7/9）。通过与前文结果进行比较，由定义 11 得到的依赖性度量 k' 不仅比较满足对知识间依赖性度量的直观认识，而且更加符合实际应用。因此，通过定义 11 可得知识依赖性程度能被较为可靠地描述。

表 1　一个简单的决策系统

U	c_1	c_2	c_3	D
x_1	0	1	1	1
x_2	1	1	1	0
x_3	1	1	1	0
x_4	0	1	1	0
x_5	0	0	2	1
x_6	0	0	2	1
x_7	1	0	2	1
x_8	1	0	2	1
x_9	0	0	2	0

3.2　经典的属性重要性度量的改进

经典的知识依赖性度量刻画属性重要性，在实际应用中往往会出现决策分类分别关于两个不同的条件分类（去掉属性 c_1 和 c_2 之后得到的 P_1 和 P_2 分别对 U 的划分）的正域中集合的元素个数相同（即 $|pos_{p_1}(D)|=|pos_{p_2}(D)|$），则经典属性重要性度量无法判断 c_1 和 c_2 属性重要性的高低，同时也使得后续应用中的结果容易出现偏差。

基于这个问题，本文考虑了条件属性子集的分类因素，即对于不同的条件属性子集，其分类程度也不同，比如分类个数。如果存在条件属性非空有限集 C 去掉一个属性 c_i 的划分为 U/P_i（其中 $P_i=C-\{c_i\}$），$U/C-\{c_i\}$ 中所含的等价类的个数越少，可以理解为 U 中的一个划分 U/P_i 中的单个等价类包含的元素越多，而根据正域的定义可知 D 关于 P_i 的正区域有变小的趋势，那么通过比较可知属性依赖度定义认为决策属性 D 关于条件属性集 P_i 的依赖性越低，因此决策属性 D 关于条件属性 c_i 的依赖度越高（c_i 的重要性越高）。

针对上述分析，可以得到如下结论：若决策分类分别关于两个不同的条件分类的正域中元素个数相等（$|pos_{P_1}(D)|=|pos_{P_2}(D)|$），则偏向于分类个数较少的那个属性集（即去掉的属性 c_i 的重要性较高）。

定义 12（可信系数）　设 $A=C\cup D$ 是属性的非空有限集合，且有 C 中去掉一个属性 c_i 的一个子集 $P=\{C-c_i\}$，那么条件属性集 C 和条件属性子集 P 对 U 的分类可分别表示为 U/C 和 U/P，记

$$\tau_{\{C-c_i\}}=\frac{|U/P|}{|U/C|}$$

为 P 关于 U/C 的可信系数。

定义 13　设 $S=(U,A,V,f)$ 是一个决策信息系统，其中 $A=U\cup D$。假设 $\varnothing\subset P\subseteq C,\varnothing\subset Q\subseteq D,U/Q\neq\{U\},p\in P$，那么 p 在 P 中关于 Q 的重要性定义为

$$sig'^{Q}_{P-\{p\}}(p)=\frac{(|S_P{}'(Q)|-\tau_{P-\{p\}}\cdot|S_{P-\{p\}}{}'(Q)|)}{|U|}$$

下面通过一个小汽车知识表示系统[1]来检验定义 13 的可行性。

表 2 为一个有关汽车的完备信息表，其中论域 $U=\{x_1,x_2,\cdots,x_8\}$，条件属性集 $C=\{$内部设计（字母 a 表示），机型（字母 b 表示），颜色（字母 c 表示）$\}$，决策属性集 $D=\{$速度（字母 d 表示），加速性能（字母 e 表示）$\}$。为了简便，将各属性值用数字代替，$V_a=\{$适中（数字 0 表示），拥挤（数字 1 表示），宽敞（数字 2 表示）$\}$，$V_b=\{$柴油（数字 0 表示），汽油（数字 1 表示），丙烷（数字 2 表示）$\}$，$V_c=\{$银色（数字 0 表示），白色（数字 1 表示），黑色（数字 2 表示）$\}$，$V_d=\{$中（数字 0 表示），高（数字 1 表示），低（数字 2 表示）$\}$，$V_e=\{$差（数字 0 表示），极好（数字 1 表示），好（数字 2 表示）$\}$。

表 2　小汽车信息决策系统

U	a	b	c	d	e
x_1	0	0	0	0	0
x_2	1	1	1	1	1
x_3	2	0	2	1	2
x_4	0	1	2	0	1
x_5	0	0	0	2	2
x_6	2	2	2	1	2
x_7	2	1	1	1	1
x_8	1	1	1	2	2

根据定义 2，有如下划分：

$U/D=\{\{x_1\},\{x_2,x_7\},\{x_3,x_6\},\{x_4\},\{x_5,x_8\}\}$

$U/C=\{\{x_1,x_5\},\{x_2,x_8\},\{x_3\},\{x_4\},\{x_6\},\{x_7\}\}$

$U/C-\{b\}=\{\{x_1,x_5\},\{x_2,x_8\},\{x_3,x_6\},\{x_4\},\{x_7\}\}$

$U/C-\{c\}=\{\{x_1,x_5\},\{x_2,x_8\},\{x_3\},\{x_4\},\{x_6\},\{x_7\}\}$

根据以上划分可求得 $|pos_{C-\{b\}}(D)|=|pos_{C-\{c\}}(D)|=4$，那么决策属性 D 与条件属性集 $P_1=\{a,c\}$ 和 $P_2=\{a,b\}$ 之间的依赖度相同，因此，根据定义 7 可知属性 b 和 c 在决策表中的重要性是相同的。然而，通过观察可知条件属性集 P_1 和 P_2 对 U 的分类中等价类个数是不同的，故可以根据定义 12 求得条件属性子集 P_1 和 P_2 关于 U/C 的可信系数分别为 $\tau_{P_1}=5/6$ 和 $\tau_{P_2}=1$，这表明利用定义 13 区分属性 b 和 c 的重要性是可行的。

另外，再分别通过变精度粗糙模型中属性的近似依赖性度量和信息熵来刻画属性重要性。结合上述几种度量方法，得到表 3 所列结果。

表 3　几种不同的度量方法比较

M	b	c
1	0	0
2	0	0
3	0	0
4	1/6	0

在表 3 中，M 表示采用的方法，M 取值为 1,2,3,4，其中 1 表示采用的方法是定义 7（经典属性重要性度量），2 表示采用的方法是属性的近似依赖性度量，3 表示采用的方法是信息熵，4 表示采用的方法是定义 13。从表中可知，当决策分类

分别关于两个不同的条件分类的正域的元素个数相等时，利用定义 13 可以比较合理地区别出二者间的重要性差异。

4 实例分析

例 1　下面给出一个 CTR 的信息决策系统[1]，其中 $U=\{x_1,x_2,\cdots,x_{14}\}$，$C=\{a_1,a_2,a_3,a_4,a_5,a_6\}$ 为条件属性集，d 为决策属性，且对象在单个条件属性 a_1,a_2,a_3,a_4,a_5,a_6 下最大的取值为 2，最小值为 1；对象在决策属性下最大的取值为 3，最小取值为 1。运用表 4 给出的决策系统来分析经典知识依赖性度量、属性近似依赖度量、信息熵以及本文提出的基于多数包含关系且加入可信系数 τ_p 的新的属性重要性度量方法之间的性能。

表 4　一个汽车测试结果决策表

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	1	1	1	1	1	1
x_2	1	1	2	1	1	2	1
x_3	1	1	2	1	1	1	1
x_4	1	2	1	1	1	1	2
x_5	1	1	2	1	1	2	1
x_6	1	1	2	2	1	2	3
x_7	1	1	2	1	1	2	3
x_8	2	2	2	2	2	1	2
x_9	1	2	2	2	2	1	1
x_{10}	1	2	2	2	2	1	1
x_{11}	2	2	2	1	2	1	2
x_{12}	2	2	2	1	1	2	2
x_{13}	1	2	2	2	1	2	1
x_{14}	2	2	1	1	2	1	2

表 4 为一个有关汽车测试的完备信息表，其中论域 $U=\{x_1,x_2,\cdots,x_{14}\}$，条件属性集 $C=\{$长度(字母 a_1 表示)，气缸数(字母 a_2 表示)，有无涡轮增压器(字母 a_3 表示)，燃料系统的类型(字母 a_4 表示)，发动机排量(字母 a_5 表示)，压缩比(字母 a_6 表示)$\}$，决策属性集 $D=\{$里程(字母 d 表示)$\}$。为了简便，将各属性值用数字代替，$V_{a_1}=\{$紧凑(数字 1 表示)，微型(数字 2 表示)$\}$，$V_{a_3}=\{$是(数字 1 表示)，否(数字 2 表示)$\}$，$V_{a_4}=\{$EFI(数字 1 表示)，2-BBL(数字 2 表示)$\}$，$V_{a_5}=\{$中等(数字 1 表示)，小(数字 2 表示)$\}$，$V_{a_6}=\{$高(数字 1 表示)，中等(数字 2 表示)$\}$，$V_d=\{$中(数字 1 表示)，高(数字 2 表示)，低(数字 3 表示)$\}$。

1)根据定义 2，有如下划分：

$U/d=\{\{x_1,x_2,x_3,x_5,x_9,x_{10},x_{13}\},\{x_4,x_8,x_{11},x_{12},x_{14}\},\{x_6,x_7\}\}$

$U/C=\{\{x_1\},\{x_2,x_5,x_7\},\{x_3\},\{x_4\},\{x_6\},\{x_8\},\{x_9,x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{13}\},\{x_{14}\}\}$

$U/C-\{a_1\}=\{\{x_1\},\{x_2,x_5,x_7\},\{x_3\},\{x_4\},\{x_6\},\{x_8,x_9,x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{13}\},\{x_{14}\}\}$

$U/C-\{a_2\}=\{\{x_1,x_4\}\},\{x_2,x_5,x_7\},\{x_3\},\{x_6,x_{13}\},\{x_8\},\{x_9,x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{14}\}\}$

$U/C-\{a_3\}=\{\{x_1,x_3\},\{x_2,x_5,x_7\},\{x_4\},\{x_6\},\{x_8\},\{x_9,x_{10}\},\{x_{11},x_{14}\},\{x_{12}\},\{x_{13}\}\}$

$U/C-\{a_4\}=\{\{x_1\},\{x_2,x_5,x_6,x_7\},\{x_3\},\{x_4\},\{x_8,x_{11}\},\{x_9,x_{10}\},\{x_{12}\},\{x_{13}\},\{x_{14}\}\}$

$U/C-\{a_5\}=\{\{x_1\},\{x_2,x_5,x_7\},\{x_3\},\{x_4\},\{x_6\},\{x_8\},\{x_9,x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{13}\},\{x_{14}\}\}$

$U/C-\{a_6\}=\{\{x_1\},\{x_2,x_3,x_5,x_7\},\{x_4\},\{x_6\},\{x_8\},\{x_9,x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{13}\},\{x_{14}\}\}$

2)利用几种经典度量方法以及定义 7，再结合上述分类信息，来进行属性重要性的刻画，得到表 5 所列结果。

表 5　各属性的重要性比较

M	a_1	a_2	a_3	a_4	a_5	a_6
1	0.214	0.285	0	0.071	0	0.071
2	0	0	0	0	0	0
3	0.124	0.180	0	0.056	0	0.022
4	0.149	0.168	0.168	0.11	0	0.084

在表 5 中，M 表示采用的方法，取值为 1,2,3,4，且与表 3 各方法相对应。在方法 1 下(经典属性重要性度量)，条件属性 a_3 和 a_5 之间以及 a_4 和 a_6 之间均无法区别它们的属性重要性，然而每组属性在实际的分类中可能出现等价类个数不同以及等价类中对象元素少部分相同或大部分相同等现象，因此，它们的重要性相等这一结果是比较缺乏说服力的。在方法 2 下(属性的近似依赖性度量)，无法区分任意条件属性的重要性，故这种方法无法直接用于实际应用。而在方法 3 下(信息熵)，条件属性 a_4 和 a_6 之间的重要性高低能够被很好地区别，但是属性 a_3 和 a_5 的重要性仍然无法区分开来；同时，条件属性集的划分越粗糙(即分类个数越少)，则该划分的初始熵越小，即该条件属性集提供给决策属性的信息量会减少，从而决策属性基于条件属性的依赖性会降低，于是可以认为决策属性 d 基于条件属性集 $P_1=\{a_1,a_2,a_4,a_5,a_6\}$ 的依赖性比基于 $P_2=\{a_1,a_2,a_3,a_4,a_6\}$ 的依赖性更低，因此属性 a_3 和 a_5 的重要性相等这一结果也是不太精确的。而通过本文提出的方法(方法 4)比较合理地区分了条件属性 a_3 和 a_5 以及 a_4 和 a_6 之间的属性重要性，故可以认为利用新的知识依赖性度量及属性重要性度量方法去解决大数据应用中的不确定性问题更有效。

结束语　针对经典知识依赖性度量在应用中存在一定的局限性，本文对它进行了改进，其策略是以多数包含关系 $c(X,Y)$ 为前提，加入可信系数 τ 来协调度量的合理性，并将改进后的知识依赖性度量方法应用于一个关于汽车检测的数据决策系统以说明它的可行性。本文的方法需要更多实例来验证，而能否将经典依赖性及属性重要性度量更好地完善还需进一步研究与探讨。

参 考 文 献

[1] Zhang Wen-xiu, Wu Wei-zhi, Liang Ji-ye, et al. Theory and method of rough set [M]. Beijing: Science Press, 2001(in Chinese)
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

[2] Hu Dan, Li Hong-xing. The measurement of dependent degree of knowledge[J]. Journal of Beijing Normal University(Natural Science), 2004, 40(3): 320-325(in Chinese)
胡丹，李洪兴. 知识依赖性的度量[J]. 北京师范大学学报(自然科学版)，2004，40(3)：320-325

[3] Lian Gang, Chen Yuan-yuan, Wu Gen-xiu. A new definition of attribute significance in consistent complete decision systems [J]. Journal of Jiangxi Normal University(Natural Science), 2008, 32(3): 326-329(in Chinese)

(下转第 306 页)

图5中，原始图像内部分人位置较为紧密，聚合在一起，仅依靠当前帧图像信息无法分割开。但聚合在一起的多人在行进过程中可能出现过分离，算法引入之前帧信息，可以将部分人分割开，有效提升了分割效果。

实验结果表明，复杂场景下，本文算法由于综合了当前帧的静态特征与前后帧的关联特征，不仅能准确地提取出前景，而且可以较好地将运动物体分割开。

结束语 本文将传统的物体分割算法同超像素分割算法结合在一起，通过对运动物体的空间关联性进行分析，提出了一种新的视频分割算法。算法使用超像素对运动前景进行表示，将视频分割转换为超像素的联接概率分析问题，并且给出了具体的求解方式。实验结果表明，文中算法在简单场景与复杂场景下均有良好的视频分割效果，较为准确地完成了运动物体的分割。今后研究工作的重点是完善运动变换特征，建立完备的运动模型，更好地建立超像素的运动关联性。

参 考 文 献

[1] Vicente S, Rother C, Kolmogorov V. Object cosegmentation[C]// 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011: 2217-2224

[2] Lee Y J, Kim J, Grauman K. Key-segments for video object segmentation[C]// 2011 IEEE International Conference on Computer Vision (ICCV). IEEE, 2011: 1995-2002

[3] Liu Y. A video object segmentation algorithm based on region selection[J]. Science Technology and Engineering, 2014, 14(6): 211-217(in Chinese)
刘毅. 一种基于区域选择的视频对象分割算法[J]. 科学技术与工程, 2014, 14(6): 211-217

[4] Trichet R, Nevatia R. Video segmentation and feature co-occurrences for activity classification[C]// 2014 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2014: 385-392

[5] Sun D, Roth S, Black M J. Secrets of optical flow estimation and their principles[C]// 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010: 2432-2439

[6] Kim K, Chalidabhongse T H, Harwood D, et al. Real-time foreground-background segmentation using codebook model[J]. Real-time imaging, 2005, 11(3): 172-185

[7] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[C]// Proceedings of the 17th International Conference on Pattern Recognition, 2004 (ICPR 2004). IEEE, 2004, 2: 28-31

[8] Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724

[9] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282

[10] http://www.sfu.ca/~ibajic/datasets.html

[11] Chen Y M, Bajic I V, Saeedi P. Moving region segmentation from compressed video using global motion estimation and Markov random fields[J]. IEEE Transactions on Multimedia, 2011, 13(3): 421-431

[12] Zeng W, Du J, Gao W, et al. Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model[J]. Real-Time Imaging, 2005, 11(4): 290-299

[13] http://sida.rdg.ac.uk/pub

(上接第276页)

连钢，程媛媛，吴根秀. 一致备决策系统中属性重要性定义的改进[J]. 江西师范大学学报(自然科学版), 2008, 32(3): 326-329

[4] Zhao Nai-gang, Li De-yu, Wang Su-ge, et al. New measure method for attribute significance in incomplete decision table and its application[J]. Computer Science, 2008, 35(8): 251-254(in Chinese)
赵乃刚，李德玉，王素格，等. 一种新的不完备决策表中属性重要性度量及其应用[J]. 计算机科学, 2008, 35(8): 251-254

[5] Zhang Wen-xiu, Xu Zong-ben, Liang Yi, et al. Inclusion Degree Theory[J]. Fuzzy Systems and Mathematics, 1996, 10(4): 1-9(in Chinese)
张文修，徐宗本，梁怡，等. 包含度理论[J]. 模糊系统与数学, 1996, 10(4): 1-9

[6] Liang Ji-ye, Xu Zong-ben, Li Yue-xiang. Inclusion degree and measures of rough setdata analysis[J]. Chinese Journal of Computers, 2001, 24(5): 544-647(in Chinese)
梁吉业，徐宗本，李月香. 包含度与粗糙集数据分析中的度量[J]. 计算机学报, 2001, 24(5): 544-647

[7] Miao Duo-qian, Wang Jue. An information representation of the concepts and operations in rough set theory[J]. Journal of Software, 1999, 10(2): 113-116(in Chinese)
苗夺谦，王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报, 1999, 10(2): 113-116

[8] Liu Cai-hui, Miao Duo-qian, Yue Xiao-dong, et al. Measures of uncertainty of knowledge and their relationships[J]. Computer Science, 2014, 41(3): 66-69(in Chinese)
刘财辉，苗夺谦，岳晓冬，等. 知识不确定性度量及其关系研究[J]. 计算机科学, 2014, 41(3): 66-69

[9] Miao Duo-qian, Hu Gui-rong. A heuristic algorithm for Reduction of knowledge [J]. Journal of Computer Research and Development, 1999, 36(6): 681-684(in Chinese)
苗夺谦，胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展, 1999, 36(6): 681-684

[10] Wang Guo-yin, He Xiao. A Self-Learning model under uncertain condition[J]. Journal of Software, 2003, 14(6): 1096-1102(in Chinese)
王国胤，何晓. 一种不确定性条件下的自主式知识学习模型[J]. 软件学报, 2003, 14(6): 1096-1102

[11] Wang Li-juan, Yang Jing-yu, Wu Chen, et al. Application of dependency relation in expanded rough set models [J]. Journal of Jiangsu University of Science and Technology (Natural Science), 2012, 26(2): 175-180(in Chinese)
王丽娟，杨静宇，吴陈，等. 扩展粗糙集模型中依赖关系及应用[J]. 江苏科技大学学报(自然科学版), 2012, 26(2): 175-180