

精简水产品安全评价指标算法研究

鄂旭^{1,2}, 杨健³, 李建革², 侯建¹, 沈德海¹

(1. 渤海大学 信息科学与技术学院, 辽宁 锦州 121001;

2. 北京交通大学 中国产业安全研究中心, 北京 100084;

3. 中航飞机起落架有限责任公司, 湖南 长沙 410200)

摘要:水产品安全评价是食品安全管理的一项重要内容,其评价体系由众多的评价指标构成。这些评价指标往往既相互依赖,又相互独立,错综复杂。因此,水产品安全评价体系中评价指标的筛选和确立就成为食品安全管理的一个重点和难点。针对这一问题,文中提出了一种精简水产品安全评价指标体系的新方法。该方法以粗糙集理论为基础,利用不可分辨关系、核属性等概念,在保持水产品安全信息系统分类能力不变的情况下,计算各评价指标重要性,选取重要评价指标,构建了简约明了的水产品安全评价体系。实例分析表明该算法是有效可行的。

关键词:食品安全;评价指标;粗糙集;核心属性

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2014)06-0223-03

doi:10.3969/j.issn.1673-629X.2014.06.055

Research on Safety Assessment Indexes Algorithm of Reducing Aquatic Product

E Xu^{1,2}, YANG Jian³, LI Jian-ge², HOU Jian¹, SHEN De-hai¹

(1. College of Information Science & Technology, Bohai University, Jinzhou 121001, China;

2. China Center for Industrial Security Research of Beijing Jiaotong University, Beijing 100084, China;

3. AVIC Landing Gear Manufacturing Corp., Changsha 410200, China)

Abstract: Aquatic product safety assessment is important contents in food safety supervision and management, and is composed of all kinds of assessment indexes which are not only dependent, but also independent. The selection and establishment of assessment indexes in aquatic product safety system is hard and difficult for food safety management. To deal with it, a new method was proposed based on rough sets by using core attributes, discernable matrix and so on. Under the condition of maintaining the classification capacity, calculate the importance of assessment indexes, select the important indexes and build the concise aquatic product safety assessment indexes system. Experimental results show that the algorithm is efficient and effective.

Key words: food safety; assessment index; rough set; core attributes

0 引言

近几年,与老百姓息息相关的食品似乎一直问题未断,从央视曝光的双汇瘦肉精事件到上海染色馒头事件,从地沟油黑幕到问题血燕的惊天骗局,从进口奶粉频现虫子到美汁源投毒案等等。这些重大食品安全问题,严重影响了人民的身心健康,同时也充分暴露出

食品安全管理体系的薄弱。食品安全问题非常复杂,它涉及从“农田”到“餐桌”的整个过程,是一个涉及多个领域、多个环节的动态问题。食品安全既受到系统内部因素的影响,如食品卫生政策、食品生产条件、食品科技水平等,又受到系统外部因素的影响,如自然灾害、环境污染等。这些因素的影响程度,需要准确测度

收稿日期:2013-06-25

修回日期:2013-09-30

网络出版时间:2014-02-11

基金项目:中国博士后基金项目(2012M520158);辽宁省百千万人才基金择优资助项目(2012921058);辽宁省教育科研项目(L2012397, L2012396, L2012400)

作者简介:鄂旭(1971-),男,博士后,教授,硕士生导师,辽宁省“百千万人才”、中国教育部学位与研究生管理中心专家,中国科技部国际科技合作计划评价专家,中国科技奖励评审专家,中国人工智能学会智能CAD与数字艺术专业委员会委员,研究方向为物联网智能计算与食品安全信息化。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140211.1448.002.html>

出来,并且需要对这些因素之间的关联程度进行准确衡量。为此,国内外学者对食品安全评价问题进行了深入的研究。目前,国内外大多数食品安全指标体系的内容是围绕导致食品不安全的主要因素来进行设计的,主要包括两个方面的内容:一个是食品中微生物污染程度,另一个是食品中有害物质含量。当前对食品安全评价的主流方法有相对评价和绝对评价相结合、排序评价和分类评价相结合、动态评价和静态评价相结合三种方法,但这些方法都需要食品安全评价的先验知识做基础,有待进一步提高^[1-3]。

粗糙集理论为此提供了有效的处理技术和方法。在 20 世纪 80 年代,波兰数学家 Pawlak 教授提出了粗糙集理论。粗糙集理论特别适合处理不确定性和模糊性数据,它不需要先验规则就能有效地分析不一致、不精确、不完整等各种不完备数据。粗糙集理论的一个重要功能就是属性约简^[4-7]。目前,国内外学者已经提出了许多种属性约简算法,例如基于可辨识矩阵的属性约简算法、基于神经网络的属性约简算法、基于遗传算法的属性约简算法、基于信息熵的属性约简算法等^[8-13]。

文中在借鉴上述方法优缺点的基础上,基于粗糙集理论提出了一种食品安全评价指标精简算法。

1 粗糙集基础知识

粗糙集的研究对象是由一个对象集合被一个多值属性集合所描述,描述符是每个对象及其属性都有的一个值。这些信息在粗糙集中用决策表来表示,对象、属性和描述符是决策表的基本组成要素。决策表可以写成是一个二维表格,表格的行表示对象,列表示对象的属性;对象信息的描述符以及类别信息相应地包含于各行中。粗糙集及其在应用中所涉及的一些基本概念如下^[4-8]。

定义 1:在粗糙集中,用四元组 $S = (U, A, V, f)$ 来表示信息系统。

其中, U 是论域,代表一个有限的非空对象集合;
 A 是一个非空有限的属性集合;

V 是属性值的集合, $V = \bigcup (V_a)$, $a \in A$, V_a 是属性 a 的值域;

f 是 $U \times A \rightarrow V$ 的一个映射函数,它的每一个取值代表研究对象的一个属性特征,即 $\forall a \in A, x_i \in U$, $f(x_i, a) \in V_a$ 。

定义 2:设信息系统 $S = (U, A, V, f)$, 设 $P \subseteq A$ 而且 $x, y \in U$, 称 x, y 关于 P 是不可分辨的,如果满足: $f(x, a) = f(y, a)$, $V_a \in P$, 由属性 P 产生的不可分辨关系也称等价关系,它将 U 划分为若干个等价类,记

为 $U/\text{IND}(P)$ 。

定义 3:令 $X \subseteq U$, 其中 R 是一等价关系。当 X 是某些等价关系 R 基本范畴的并集时,就称 X 为 R 可定义的,否则 X 是 R 不可定义的。 R 的可定义集是论域 U 的子集,在知识库 K 中可被精确的定义,而在知识库中 R 的不可定义集则不能被定义。 R 的可定义集被称作 R 精确集,而 R 的不可定义集被称作非精确集或 R 粗集。

定义 4:假定 $K = (U, R)$ 代表知识库,对于每个子集 $X \in U$ 和一个等价关系 $R \in \text{IND}(K)$,可以根据等价关系 R 的基本集合的描述对集合 X 进行划分。为了判定 $\{\text{des}(Y_i), Y_i \in R\}$ 精确的说明 X 中对象的隶属度情况,采用两个子集:

$$R_-(X) = \bigcup \{Y \in U/R; Y \subseteq X\}$$

$$R^-(X) = \bigcup \{Y \in U/R; Y \cap X \neq \emptyset\}$$

分别表示 X 的 R 下近似和 R 上近似。

它们也可以表征如下:

$$R_-(X) = \{x \in U; [x]R \subseteq X\}$$

$$R^-(X) = \{x \in U; [x]R \cap X \neq \emptyset\}$$

$\text{pos}_R(X) = R_-(X)$ 表示 X 在 R 上的正域, $\text{neg}_R(X) = U - R_-(X)$ 代表 X 在 R 上的负域,把 $\text{bn}_R(X)$ 代表 X 的取值边界。这几个概念在粗糙集理论中非常重要,在有关属性和属性值的约简中都起到了重要作用。集合 X 的上近似、下近似概念可用一个二维近似空间中的图例来描述。该空间是用划分成基本区域的长方块构成的 (U, R) 来定义的,每个基本长方块区域表示 R 的一个等价类,阴影区域表示 X 的边界,表示 X 的不可确定区域,不包括 $R_-(X)$ 的其余所有的区域是负域。

定义 5:假定 $Q \subseteq P$ 独立,且存在 $\text{IND}(Q) = \text{IND}(P)$,则可以把 Q 看作是关系簇集 P 的一个约简。在 P 中所有绝对必要的关系集合称为 P 的核,记为 $\text{core}(P)$,其结果为 P 的所有约简集合的交集,即: $\text{core}(P) = \bigcap \text{red}(P)$ 。

定义 6:设系统 $S = (U, A, V, f)$, 其中 $A = C \cup D$, $a(x)$ 是 x 在属性 a 上的值,分辨矩阵 M 为:

$$(C_{ij}) = \begin{cases} a \in A: & a(x_i) \neq a(x_j) \wedge D(x_i) \neq D(x_j) \\ 0 = \emptyset & D(x_i) = D(x_j) \\ -1 & a(x_i) = a(x_j) \wedge D(x_i) \neq D(x_j) \end{cases}$$

同时分辨矩阵中的核就是组合数为 1 的属性。

定义 7:属性的重要性。在分辨矩阵 $M = (C_{ij})_{N \times N}$ 中,相应的属性重要性计算公式为:

$$f(a) = \sum_{i=1}^n \sum_{j=1}^n \frac{\lambda_{ij}}{t_{ij}}$$

其中, $\lambda_{ij} = \begin{cases} 0 & a \notin c_{ij} \\ 1 & a \in c_{ij} \end{cases}$; t_{ij} 表示 c_{ij} 中包含的属性
的个数。

2 算法描述与实例分析

该算法主要是通过数据自身的不可分辨关系来实现的,它的基本思想就是从数据间的不可分辨关系对论域进行划分,用一对近似集合对给定的概念进行近似表示。给定对象间的一个等价关系,用不可分辨对象类所形成的上近似集和下近似集来表示。上近似集表示确定属于给定类别的最大的对象集合,而下近似集则表示可能属于给定类别的最小的对象集合。

2.1 算法描述

输入: $(U, A \cup \{d\})$ 代表决策表, $A = \cup a_i, i = 1, \dots, n$ 。

输出: 决策表的属性约简集合 R 。

步骤:

- (1) 将条件属性集合作为约简后得到的属性集合,即 $\text{reduct} = R$;
- (2) 计算决策表的可辨识矩阵 M , 并将所有不包含核属性的属性组合 S 找出;
- (3) 将所有不包含核属性的属性组合 S 用析取范式的形式表示:

$$P = \bigwedge \{ \bigvee a_{ik} : i = 1, \dots, s, k = 1, \dots, m \} ;$$

- (4) 应用属性重要性公式计算属性重要性;
- (5) 排序属性重要性,选择最小属性值 a , 并计算 $\text{reduct} = \text{reduct} - \{a\}$;
- (6) 对属性集合作约简操作判断: 满足判定条件, 则删除冗余属性及矛盾样本, 并 $i = i + 1$, 跳回(5); 不满足判定条件, 则整理约简属性集合, 结束操作。

2.2 实例分析

给定一个水产品安全信息表,如表 1 所示。其中, a, b, c 为水产品安全信息表中的条件属性, d 为决策属性, 即 $C = \{a, b, c\}, D = \{d\}$ 。

表 1 水产品安全信息表

U	a	b	c	d
1	0.9	2	1	1
2	1.1	0.8	1	0
3	1.3	3	2	0
4	1.4	1	1	1
5	1.4	2	1	1
6	1.2	1	1	1
7	1.8	3	2	1
8	4	3	2	1

根据提出的属性约简算法进行等价划分:

$$U \mid \text{IND}(a, b, c) = \{ \{1\}, \{2\}, \dots, \{8\} \};$$
$$U \mid \text{IND}(a) = \{ \{1\}, \{2\}, \{3\}, \{4, 5\}, \{7\}, \{8\} \};$$
$$U \mid \text{IND}(b) = \{ \{1, 5\}, \{2\}, \{3, 7, 8\}, \{4, 6\} \};$$
$$U \mid \text{IND}(c) = \{ \{1, 2, 4, 5, 6\}, \{3, 7, 8\} \}.$$

由分类质量公式得:

$$r_c(D) = \text{card}(\text{POS}_c(D)) / \text{card}(U) = 8/8 = 1;$$
$$r_{C \setminus \{a\}}(D) = \text{card}(\text{POS}_{C \setminus \{a\}}(D)) / \text{card}(U) = 3/8;$$
$$r_{C \setminus \{b\}}(D) = \text{card}(\text{POS}_{C \setminus \{b\}}(D)) / \text{card}(U) = 5/8;$$
$$r_{C \setminus \{c\}}(D) = \text{card}(\text{POS}_{C \setminus \{c\}}(D)) / \text{card}(U) = 8/8.$$

由属性重要性公式得:

属性 a 的重要性为 $r_c(D) - r_{C \setminus \{a\}}(D) = 1 - 0.375 = 0.625$;

属性 b 的重要性为 $r_c(D) - r_{C \setminus \{b\}}(D) = 1 - 0.625 = 0.375$;

属性 c 的重要性为 $r_c(D) - r_{C \setminus \{c\}}(D) = 1 - 1 = 0$ 。

所以,属性重要性降序排序为 $\{a, b, c\}$ 。设初始重要属性集为 $R = \{\emptyset\}$, 选取重要性最大的属性 a , 即 $R = \{a\}$, 计算出 $r_R(D) = 0.6 \neq r_c(D)$, 故再选取属性 b , 则 $R = R \cup \{b\} = \{a, b\}$, 再计算 $r_R(D) = 1$, 即此时 $r_R(D) = r_c(D)$, 所以 $R = \{a, b\}$ 就是所求的一个最小相对约简集。

3 结束语

文中针对水产品安全评价指标的选定与精简问题进行了研究,利用可分辨矩阵和正域概念提出了一种精简水产品安全评价指标的新方法。该方法不但能够准确选定水产品安全评价指标,而且能够在保持原有信息系统分类能力不变的情况下精简水产品安全评价指标,发现数据中蕴涵的知识模式。但该算法也存在算法时间复杂度较高等缺点,有待于以后进一步研究。

参考文献:

[1] 邓聪文,朱雪冬,王俊能. 食品安全评价及其方法简述[J]. 畜禽业,2009(6):8-10.

[2] 郭旭强,王在建,王秀霞,等. 影响水产食品安全因素的分析[J]. 齐鲁渔业,2009,26(12):49-51.

[3] 鄂旭,韩芳,侯建,等. 面向食品安全评价的属性约简方法研究[J]. 吉林大学学报(信息科学版),2013,31(3):314-319.

[4] Pawlak Z. Rough sets and fuzzy sets[J]. Fuzzy Sets and Systems,1985,17:99-102.

[5] Krysikiewicz M. Rough set approach to incomplete information system[J]. Information Sciences,1998,112:39-49.

[6] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社,2005.

[7] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北

度上文中方法相比新建簇方法稍高,但相比合并簇方法低很多。

2)单位信道节点数比较。

如图6所示,文中方法一方面在簇处理之前先对现有簇进行簇分解判决,分解成员数量低于一定值的簇,并释放该簇使用的 ZigBee 信道;另一方面,在新建簇时进行建簇条件判决,禁止孤儿节点个数较少也成簇的情况。避免了单位 ZigBee 信道节点数少的簇出现。

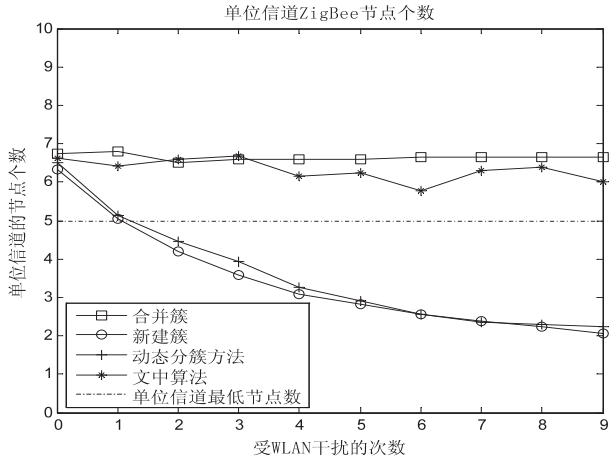


图6 单位信道 ZigBee 节点数比较

通过仿真比较,文中提出的算法时间复杂度稍高于新建簇方法,但可以很好地解决剩余孤儿节点以及吞吐量问题,通过参数设置可以控制整个网络的单位 ZigBee 信道的节点数,提高网络的吞吐量,解决 WLAN 干扰问题。

4 结束语

为了解决 ZigBee 网络中簇受 WLAN 干扰的问题,文中提出了一种合并簇、新建簇和分解簇联合使用的抗 WLAN 干扰动态分簇策略。可以在解决 WLAN 干扰问题的前提下,稍微增加时间复杂度,大大提高网络吞吐量。分解簇方法和新建簇条件判决方法,对现有簇和新建簇单位 ZigBee 信道节点数进行控制,确保整个网络的单位 ZigBee 信道节点数不低于某个值,对优化整个 ZigBee 网络吞吐量有着很大的作用。

参考文献:

[1] 潘伟,黄东. 基于 ZigBee 技术的无线传感网络研究[J]. 计算机技术与发展,2008,18(9):244-247.

[2] 李蛟,杨仁锬,肖峻. 2.4GHz 无线技术标准及 ZigBee 抗干扰性能[J]. 电信工程技术与标准化,2006,19(3):31-35.

[3] 刘荣伍,蒋挺,周正. 基于 UWB、ZigBee 的 WPAN 与 WLAN 的干扰[J]. 无线电通信技术,2006,32(3):54-56.

[4] 姜伟,朱凯,刘童. 浅谈 ZigBee 和 Wi-Fi 的共存和干扰[J]. 科技视界,2013(16):43-43.

[5] IEEE 802.15.1 Specification. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs) [S]. 2006.

[6] 王鸿运,应自炉,肖伟明,等. ZigBee 和 Wi-Fi 组成无线网关的干扰及共存机制[J]. 工业控制计算机,2012,25(12):77-78.

[7] 陈聪,冯玉林,施惠昌. ZigBee 和 Wi-Fi 的干扰和共存[J]. 计算机工程与设计,2006,27(18):3397-3399.

[8] 王海涛. 无线传感网络中的分簇算法综述[J]. 传感器世界,2011,17(4):6-10.

[9] Yu J Y, Chong P H J. A survey of clustering scheme for mobile ad hoc networks[J]. IEEE Communications Surveys and Tutorials, 2005, 7(1):32-48.

[10] 刘瑞霞,李春杰,郭强,等. 基于 ZigBee 网状网络的分簇路由协议[J]. 计算机工程,2009,35(3):161-163.

[11] Kang M S, Chong J W, Hyun H, et al. Adaptive interference-aware multi-channel clustering algorithm in a ZigBee network in the presence of WLAN interference[C]//Proc of 2nd international symposium on wireless pervasive computing. San Juan:[s. n.], 2007.

[12] Ko D N, Chong J W, Jung B H, et al. Dynamic clustering schemes of ZigBee networks in the presence of WLAN interference[C]//Proc of IEEE military communications conference. San Diego, CA: IEEE, 2008:1-7.

[13] Park T R, Kim T H, Choi J Y, et al. Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA/CA[J]. Electronics Letter, 2005, 41(18):1017-1019.

[14] Won C, Youn Jong-Hoon, Ali H, et al. Adaptive radio channel allocation for supporting coexistence of 802.15.4 and 802.11b[C]//Proc of VTC2005. [s. l.]: [s. n.], 2005:2522-2526.

[11] Hu X H, Cercone N. Learning in relational databases: A rough set approach[J]. Computational Intelligence, 1995, 11(2):323-337.

[12] 常犁云,王国胤,吴渝. 一种基于 Rough set 理论的属性约简及规则提取方法[J]. 软件学报,1999,10(11):1206-1211.

[13] 鄂旭,高学东,喻斌. 基于扫描向量的属性约简方法[J]. 北京科技大学学报,2006,28(6):604-608.

(上接第 225 页)

京:科学出版社,2006.

[8] 曾黄麟. 粗糙集理论及其应用[M]. 重庆:重庆大学出版社,1996.

[9] E Xu, Yang Yuqiang, Ren Yongchang. A new method of attribute reduction based on information quantity in an incomplete system[J]. Journal of Software, 2012, 7(8):1881-1888.

[10] 武森,高学东, Bastian M. 数据仓库与数据挖掘[M]. 北京:冶金工业出版社,2003.

作者：[鄂旭](#)，[杨健](#)，[李建革](#)，[侯建](#)，[沈德海](#)，[E Xu](#)，[YANG Jian](#)，[LI Jian-ge](#)，[HOU Jian](#)，[SHEN De-hai](#)

作者单位：[鄂旭, E Xu\(渤海大学 信息科学与技术学院, 辽宁 锦州 121001; 北京交通大学 中国产业安全研究中心, 北京 100084\)](#)，[杨健, YANG Jian\(中航飞机起落架有限责任公司, 湖南 长沙, 410200\)](#)，[李建革, LI Jian-ge\(北京交通大学 中国产业安全研究中心, 北京, 100084\)](#)，[侯建, 沈德海, HOU Jian, SHEN De-hai\(渤海大学 信息科学与技术学院, 辽宁 锦州, 121001\)](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2014(6)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201406055.aspx