数据挖掘中粗糙决策规则及其不确定性研究

邓玮舛,余永权
(广东工业大学 计算机学院,广东 广州 510090)

摘 要:数据分析中产生的粗糙决策规则通常具有不确定性,需要适当的不确定性量度。借鉴变精度粗糙集理论思想,讨论了几种粗糙决策规则量度方法,采用基于信息熵的方法给出了变精度粗糙集意义下基于修正信息熵的不确定性量度函数,兼顾到了规则不确定性的两个方面:一致性和随机性,还能有效处理噪声对数据一致性的影响,对"几乎一致性规则"有保护作用。通过举例比较了 γ_0、H^{det} 和 H^{VPRS},结果表明 H^{VPRS} 更适合于评价从有噪声数据中提取的粗糙决策规则。

关键词:数据挖掘;粗糙集;变精度粗糙集;粗糙决策规则;不确定性量度
中图分类号:TP18 **文献标识码**:A **文章编号**:1673-629X(2008)08-0050-04

Rough Decision Rules and Its Uncertainty Research in Data Mining

DENG Wei-chuan, YU Yong-quan
(Faculty of Computer, Guangdong University of Technology, Guangzhou 510090, China)

Abstract: The evaluation of the uncertainty of rough decision rules in data analyzing needs proper uncertainty measures. Several methods of measuring rough decision rules are discussed and an information entropy-based uncertainty measures are presented based on variable precision rough set theory, which can deal with the two aspects of uncertainty of rules, namely inconsistency and randomness. Also they consider the influence of the noise in the data upon the consistency of the rules and can propose "nearly consistent rules". A simple example is used to compare γ_0, H^{det} and H^{VPRS}, illustrating that H^{VPRS} is better for evaluating rough decision rules.

Key words: data mining; rough sets; variable precision rough set; rough decision rules; uncertainty measure

0 引 言

粗糙集(Rough sets)理论是波兰数学家 Z. Pawlak 在1982年提出的一种分析数据的数学理论[1]。该理论在分类的意义下定义了模糊性和不确定性的概念,是一种处理不确定、不相容数据和不精确问题的新型数学工具,已经获得越来越广泛的关注。粗糙集理论的基本思想是,根据已知数据自身的不可分辨关系,通过一对近似集合(下近似集合和上近似集合),对某一给定概念进行近似表示。在粗糙集数据分析中,可以通过这种近似形成一组粗糙决策规则,以对新数据进行分类和预测[2]。在粗糙集理论中,通常以近似度 γ_0 来衡量规则集合的不确定性,其缺点是只考虑了规则集合的一致性,而忽略了由划分的粒度引起的随机性。文献[3]中指出了 γ_0 的局限性,并给出了基于信息熵的粗糙集数据分析不确定测量函数。文中在此基础上,结合变精度粗糙集理论的思想,给出了基于修正信息熵的不确定性测量函数。

收稿日期:2007-11-05
基金项目:国家自然科学基金资助项目(60272089)
作者简介:邓玮舛(1984-),女,江西赣州人,硕士研究生,研究方向为智能工程与软计算、粗糙集;余永权,教授,博导,研究方向为智能工程与软计算。

1 基本概念

1.1 粗糙集

在粗糙集理论中,研究的领域称为论域,对论域的全部知识均来自对通过观察、测试等手段获得的数据的划分,为了数据的分析方便,通常将数据表示为二维数据表的形式[1]。

令 $S=(U,A,V_a,f_a)_{a\in A}$ 为一个知识表达系统,其中,U 表示对象的非空有限集合,称为论域;A 为非空、有限的属性集合;对每个属性 $a\in A$,V_a 是属性 a 的值域;$f_a: U\to V_a$ 为从论域 U 到属性 a 的值域的映射。

令 $T=(U,C,d,V_a,f_a)_{a\in C\cup d}$ 为一决策表,若 $A=C\cup d$,且 $S=(U,A,V_a,f_a)_{a\in A}$ 为一知识表示系统。其中,$C,d\subset A$ 分别称为条件属性集和决策属性集。

对任意的 $\varnothing\neq R\subseteq A$,有二元关系 $\mathrm{IND}(R)=\{(x,y)\in U^2 \mid \forall a\in R, f_a(x)=f_a(y)\}$,称 $\mathrm{IND}(R)$

为不可分辨关系,也称等价关系。它把 U 划分为有限个集合,称为等价类。在每个等价集合中,对象间是不可分辨的。对于 $\forall x \in U$,它的 P 等价类定义为 $[x]_P = \{y \in U \mid (x,y) \in \mathrm{IND}(P)\}$,称为 P 基本知识或基本集合,记为 U/P。

粗糙集理论认为,对被研究的对象集合 U,可以从中获得信息的精细或粗糙程度,即粒度,它是通过不可分辨关系对 U 的划分 P 来描述的。粗糙集可以用两个精确集,即粗糙集的下近似和上近似来描述。

对于 $X \subseteq U$ 和 U 上的不可分辨关系 R,定义[1]:

$$\underline{R}(X) = \bigcup \{Y \subseteq U/R \mid Y \subseteq X\}$$

$$\overline{R}(X) = \bigcup \{Y \subseteq U/R \mid Y \cap X \neq \emptyset\}$$

分别为 X 的 R 下近似集合和 X 的 R 上近似集合,称 $< \underline{R}(X), \overline{R}(X) >$ 为 X 的 R 粗糙集。

对于 $X \subseteq U$ 和给定的 R,若 $x \in \underline{R}(X)$,则 x 确定地属于 X,若 $x \notin \overline{R}(X)$,则 x 确定地不属于 X,若 $x \in \overline{R}(X) \setminus \underline{R}(X)$,则 x 可能属于 X,也可能不属于 X。故称 $\underline{R}(X)$ 为 X 的正域,$U \setminus \overline{R}(X)$ 为 X 的负域,$\overline{R}(X) \setminus \underline{R}(X)$ 为边界[1]。

决策表可由其上的不可分辨关系生成一系列粗糙决策规则。设 $Q \subseteq C, X \in U/Q, Y \in U/d$,定义 $Q \to d \subseteq U/Q \times U/d$ 为由 Q 到 d 的粗糙规则集合,有 $< X, Y > \in Q \to d \Leftrightarrow X \subseteq \underline{Y}_Q$。记 $Q \xrightarrow{\det} d = \{< X, Y > \in Q \to d \mid X \subseteq \underline{Y}_Q\}$ 为一致性规则集合。定义 $Q \to d$ 的正域为:$V_0 = \bigcup \{X \in U/Q : \exists Y \in U/d, < X, Y > \in Q \xrightarrow{\det} d\}$,由此,可定义 $Q \to d$ 的近似度 $\gamma = \frac{|V_0|}{|U|}$,其中,$|\cdot|$ 表示集合包含的元素个数[1]。

1.2 变精度粗糙集

变精度粗糙集[4]是对标准粗糙集理论的一种扩展。它通过设置阈值参数,放松了标准粗糙集理论对近似边界的严格定义。

给定决策表 T 和阈值 $0.5 < \beta \leqslant 1$,定义[4]:

$$\underline{R_\beta}(X) = \bigcup \{[x]_R : \frac{|[x]_R \cap X|}{|[x]_R|} \geqslant \beta\}$$

$$\overline{R_\beta}(X) = \bigcup \{[x]_R : \frac{|[x]_R \cap X|}{|[x]_R|} > 1 - \beta\}$$

分别为 X 的 R 下 β 近似和 X 的 R 上 β 近似。

当 $\beta = 1$ 时,变精度粗糙集即为标准粗糙集。随着 β 减小,变精度粗糙集的近似边界区域变窄,即变精度粗糙集意义下的不确定区域变小。因此,变精度粗糙集对数据不一致性有一定的容忍度,在某些场合可以增强产生规则的鲁棒性,提高预测精度。

类似地,可以定义 β 粗糙规则集 $Q \xrightarrow{\beta} d \subseteq U/Q \times U/d$,有 $< X, Y > \in Q \xrightarrow{\beta} d \Leftrightarrow X \subseteq \underline{Y^\beta}$ 记为 $Q \xrightarrow{\beta - \det} d = \{< X, Y > \in Q \xrightarrow{\beta} d : X \subseteq \underline{Y^\beta}\}$ 为 β 一致性规则。$Q \xrightarrow{\beta} d$ 的正域为 $V_1 = \bigcup \{X \in U/Q : < X, Y > \in Q \xrightarrow{\beta - \det} d\}$。

1.3 信息熵及其划分粒度

在信息论中,熵可以用来度量信息系统的不确定性。对于定义在给定论域上的随机变量 X,设其概率分布为 $\{P_1, \cdots, P_s\}$,$\sum_{i=1}^{s} P_i = 1$,则 X 的信息熵为[3]:

$$H(X) = \sum_{i=1}^{s} p_i \log_2 \frac{1}{p_i}$$

对于论域 U 上的任一不可区分关系 R,可以看作是 U 上的一个随机变量,其取值为 R 对 U 的划分 U/R 所形成的不可区分类 $\{X_1, X_2, \cdots, X_s\}$,则其概率分布可估计为:

$\hat{p}(X = X_i) = \frac{|X_I|}{|U|}, i = 1,2,\cdots,s$;于是,可定义划分 U/R 的熵 $H(X) = \sum_{i=1}^{s} \frac{|X_i|}{|U|} \log_2 \frac{|U|}{|X_i|}$。

$H(X)$ 所估计的是在不可区分关系对论域的划分下,对论域中任一数据对象,要确定其属于哪个不可区分类所需最少次数[3],因而可以用来量度不可区分关系 R 对论域划分的粒度。由 $U/R = \{X_1, X_2, \cdots, X_s\}$,若 $s = 1$,则 $X_1 = U, \hat{p}(X = X_i) = 1/|U|, i = 1,2,\cdots,s$,这是对论域最粗的划分。此时,对于 $\forall x \in U$,都有 $x \in X_1$,该结果没有产生任何信息增益。若 $s = |U|$,则 $|X_1| = 1, \hat{p}(X = X_i) = 1/|U|, i = 1,2,\cdots,s$,这是对论域最细的划分。此时,对于 $\forall x \in U$,难以确定其属于哪个不可区分类。该结果产生了最大的信息增益,其值等于 $\log_2 |U|$。

2 粗糙决策规则的几种不确定性量度

在粗糙集数据分析中,粗糙决策规则既包含一致性规则又包含不一致性规则,从数据中发现这些规则并进行准确分类和预测是粗糙集数据分析的主要目标[2,3]。粗糙决策规则是通过不可区分关系对论域的划分和对决策类的粗糙近似产生的。对于一个决策表,选择其上不同的不可区分关系,可以形成不同的粗糙决策规则集合。对于不同规则集合,定义合理的不确定性量度,可以作为决策模型选择的标准。

2.1 以近似度因子作为粗糙决策规则不确定性量度

在粗糙集理论中,通常用近似度 γ_0 来衡量规则集合的不确定性。近似度 γ_0 为一致性规则对应的数据对象在全部数据对象中所占的比重,是粗糙规则集合总

体一致性的量度。若 $\gamma_0 = 1$,则所有规则都是一致的,γ_0 越小,说明规则集合的随机性越大[5]。

对于 U 上的两个不可区分关系 B_1、B_2 及其对 U 的划分 U/B_1,U/B_2,设形成的粗糙决策规则集合的近似度 γ,分别为 γ_1 和 γ_2。若 U/B_1 为较 U/B_2 更精细的划分,即 $U/B_1 \prec U/B_2$,则 $\gamma_1 \geqslant \gamma_2$[5]。

可见,对论域更精细的划分可以提高规则集合的一致性。但从另方面说,划分越精细,不可区分类的粒度就越小。粒度过小,γ_0 虽然可能很高,但每条规则基于的对象太少,已知数据对规则的支持数不够,使规则产生的随机性增大,缺乏对数据的代表性,从而影响其对新数据对象的分类预测能力。因此,以 γ_0 作为粗糙决策规则不确定性的量度,仅反映了规则集合的一致性,忽略了随机性,所以不能完全描述粗糙规则集合的不确定性。尤其是当两个不同的不可区分关系生成的两组粗糙规则具有相同的 γ_0 值时,无法分辨其优劣。

2.2 粗糙集意义下基于信息熵的规则不确定性量度

由于信息熵可以量度对论域划分的粒度,因此,可将粗糙决策规则集合也看作是对论域的一个划分,应用信息熵来测量规则集合的不确定性。标准粗糙集理论认为,当选择了某个条件属性子集 B,则对论域的全部知识均只来自 B 对论域 U 的划分,粗糙决策规则集合 $B \to d$ 来自于 U/B 对 U/d 的近似。因此,只有 $B \to d$ 的正域 V_0 是已知的,而边界区域 $U \setminus V_0$ 是未知的。由此假定,只有正域 V_0 中的对象有确定的 B 到 d 的对应关系,而边界区域 $U \setminus V_0$ 中的对象,则被看作是随机分配。因此,$U \setminus V_0$ 中的每个对象均各自对应一条随机规则。Düntsch I. 等构造了满足上述假定的不可区分关系 $I^{\det}$,并给出了 $I^{\det}$ 对论域的划分 $U/I^{\det}$,同时给出了粗糙决策规则集合不确定性的信息熵量度 $H^{\det}$,对一致性规则和不一致性规则区别对待[3]。

给定论域 U 及 $C \cup d = A$,$B \subseteq C$,设 B 和 d 对论域 U 的划分分别为:$U/B = \{X_1, X_2, \cdots, X_s\}$ 和 $U/d = \{Y_1, Y_2, \cdots, Y_t\}$,有 $V_0 = X_1 \cup \cdots \cup X_c$,$C \leqslant S$。定义不可区分关系[3] $I^{\det} = \{(x,y) \in U^2 \mid x = y \vee \exists i \leqslant c, x, y \in X_i\}$;$I^{\det}$ 对论域的划分形成的不可区分类分为两部分:一是构成 V_0 的 $X_i (i \leqslant c)$;二是 $U \setminus V_0$ 中每一个元素 x 均构成一个不可区分类。于是可以得到各个不可区分类的概率分布的估计:

$$\hat{\varphi}_i = \begin{cases} \dfrac{|X_i|}{|U|}, & \text{若 } i \leqslant c \\ \dfrac{1}{|U|}, & \text{其它} \end{cases}$$

标准粗糙集意义下基于信息熵的不确定性量度 $H^{\det}(B \to d) = H(I^{\det}) = \sum_i \hat{\varphi}_i \log_2 \dfrac{1}{\hat{\varphi}_i}$;不难看出[5],$I^{\det} \subseteq I_d$,$U/I^{\det} \prec U/I^d$。

结合上面两式得到:$H^{\det}(B \to d) = \sum_{i \leqslant c} \dfrac{|X_i|}{|U|} \log_2 \dfrac{|U|}{|X_i|} + \dfrac{|U \setminus V_0|}{|U|} \log_2 |U|$;令 $H^{\text{con}}(B \to d) = \sum_{i \leqslant c} \dfrac{|X_i|}{|U|} \log_2 \dfrac{|U|}{|X_i|}$,$H^{\text{inc}}(B \to d) = \dfrac{|U \setminus V_0|}{|U|} \log_2 |U|$,则 $H^{\det} = H^{\text{con}} + H^{\text{inc}}$,$H^{\text{con}}$ 表示一致规则的不确定度,H^{inc} 表示不一致规则的不确定度。$H^{\det}$ 完全基于标准粗糙集意义,认为仅有一致规则,也即正域的概率分布有意义,而不考虑边界区域。这样,对于一致规则,粒度越大,即 $|X_i|$ 越大,H^{con} 的值越小。由于决策类 U/I^d 是固定不变的,所以,粒度越大,则规则的支持数越大,随机性越小。可见,H^{con} 可以度量一致性规则的随机性;而对于所有不一致规则,假定支持规则的每个数据对象对应一条随机规则,从而 H^{inc} 包含了不一致规则在信息熵意义下的最大不确定度,通过给不一致规则赋以最大的随机性来对规则的不一致性进行量度。所以说,$H^{\det}$ 在一定程度上兼顾了粗糙规则集合不确定性的两个方面。

2.3 变精度粗糙集意义下基于信息熵的规则不确定性量度

$H^{\det}$ 的原则是将规则集合的正域和边界区域截然分开,因此可能造成对弱不一致性的夸大。为了恰当处理数据噪声引起的规则不一致性问题,需要对规则集合的边界重新理解和定义,受 VPRS 模型的启发,下面构造了一个基于 VPRS 的信息熵量度,来处理噪声引起的不一致性。

给定论域 U 及 $C \cup d = A$,$Q \subseteq C$,设 Q 和 d 对论域 U 的划分分别为:$U/Q = \{X_1, X_2, \cdots, X_S\}$ 和 $U/d = \{Y_1, Y_2, \cdots, Y_t\}$,且有 $0.5 < \beta \leqslant 1$,$c + b \leqslant s$,使得 $V_0 = X_1 \cup \cdots \cup X_c$,$V_1 = V_0 + X_{c+1} \cup \cdots \cup X_{c+b}$。定义不可区分关系[4]:$I^{\text{VPRS}} = \{(x,y) \in U \times U \mid x = y \vee \exists i \leqslant c + b, x, y \in X_i\}$

I^{VPRS} 对论域的划分形成的不可区分类为两部分:一部分是构成 V_1 的 $X_i (i \leqslant c)$;另一部分则是 $U \setminus V_1$ 中每一个元素 x 均构成一个不可区分类。这样,通过用基于 VPRS 的粗糙规则的正域 V_1 来替代标准粗糙集的正域 V_0,放宽了一致规则的定义,将一致度(准确度)不低于 β 的不一致规则等同于一致规则。从而得到:$U/I^{\text{VPRS}} = \{X_i \mid i \leqslant c + b\} \cup \{\{x\} \mid x \in X_j, c + b < j \leqslant s\}$,于是可以得到各个不可区分类的概率分布的估计[4~6]:

$$\hat{\varphi}_i = \begin{cases} \dfrac{|X_i|}{|U|}, & \text{若 } i \leqslant c + b \\ \dfrac{1}{|U|}, & \text{其它} \end{cases}$$

变精度粗糙集意义下基于信息熵的不确定性量度[4~6]:

$$H^{VPRS}(Q\xrightarrow{\beta}d)=H(I^{VPRS})=\sum_i \hat{\varphi}_i\log_2\frac{1}{\hat{\psi}_i}$$

结合上面两式子:$H^{VPRS}(Q\to d)=\sum_{i\leqslant c}\frac{|X_i|}{|U|}\log_2\frac{|U|}{|X_i|}+\sum_{c<i\leqslant c+b}\frac{|X_i|}{|U|}\log_2\frac{|U|}{|X_i|}+\frac{|U\setminus V_0|}{|U|}\log_2|U|$,令 $H^{con}(Q\to d)=\sum_{i\leqslant c}\frac{|X_i|}{|U|}\log_2\frac{|U|}{|X_i|}$,$H^{nco}(Q\to d)=\sum_{c<i\leqslant c+b}\frac{|X_i|}{|U|}\log_2\frac{|U|}{|X_i|}$;$H^{inc}(Q\to d)=\frac{|U\setminus V_0|}{|U|}\log_2|U|$,则 $H^{VPRS}=H^{con}+H^{nco}+H^{inc}$。这样,将不确定性量度分为三部分:一致规则的不确定度,几乎一致规则的不确定度和不一致规则的不确定度。上式中的 H^{con} 用来量度标准粗糙集意义下的一致性规则的随机性,H^{nco} 用来量度几乎一致性规则的随机性。对不一致性较强的规则将视为不一致规则,通过给这些不一致规则赋以最大的随机性来量度它们的不一致性,即认为上式中的每个对象均各自对应一条随机规则,因此,上式中包含了不一致规则在信息熵意义下的最大不确定度。

H^{VPRS} 以 VPRS模型的边界定义为依据,将几乎一致性规则与一致性规则等同看待,起到了保护几乎一致规则的作用,从而可以有效地处理噪声对数据一致性的影响。用 H^{VPRS} 作为属性选取的评判标准来求取属性约简可使获得的规则具有一定的噪声容忍度。

3　粗糙决策规则不确定性量度比较分析

本节将通过构造一个简单例子来进一步比较近似度 γ_0 和两种粗糙集模型下信息熵量度 H^{det}、H^{VPRS}。

已知决策表 $T=(U,C,d,V_a,f_a)_{a\in C\cup d}$,$|U|=200$,$U/d=\{D_1,D_2\}$,有 5 个不完全相同的条件属性子集 $Q_i\subseteq C$,$i=1,\cdots,5$,各个 Q_i 对 U 的划分

$U/Q_1=\{B_1,B_2,B_3,B_4\}$;$U/Q_2=\{T_1,T_2,T_3,T_4\}$;$U/Q_3=\{X_1,X_2,X_3,X_4\}$;$U/Q_4=\{Y_1,Y_2,Y_3,Y_4\}$;$U/Q_5=\{Z_1,Z_2,Z_3,Z_4,Z_5,Z_6,Z_7,Z_8\}$

$Q_i\to d$ 的分布如表1所示。设 $\beta=0.9$,计算 $Q_i\to d$ 的 γ_0,H^{det},H^{VPRS},如表 2 所示。

(1)标准粗糙集的近似度 γ_0 量度的是一致性规则在规则集中所占的比重。$Q_4\to d$ 中全为一致规则,故 $\gamma_0=1$;$Q_1\to d$ 和 $Q_2\to d$ 中全为不一致规则,故 $\gamma_0=0$。

(2)但 γ_0 不能反映规则集合的随机性,如 $Q_3\to d$ 和 $Q_5\to d$ 具有相同的 γ_0 值,却无法区别两个规则集的好坏。从表中可明显看出 $Q_3\to d$ 中规则粒度大,随机性小。这一点从 H^{det} 的取值中有明显体现,$Q_3\to d$ 的 H^{det} 取值比 $Q_5\to d$ 的小,说明前者随机性小,进而可判断规则集 $Q_3\to d$ 要好些。

(3)H^{det} 可以兼顾规则集合的不一致性和随机性两个方面,但有时可能夸大由噪声而引起的微弱不一致性,将微弱不一致规则与不一致规则等同对待,如 $Q_1\to d$ 和 $Q_2\to d$ 的 H^{det} 相同,但 $Q_2\to d$ 明显比 $Q_1\to d$ 规则集要好,此时从 H^{det} 的取值无法区分二者的优劣,H^{VPRS} 则很好地体现了这一点。

(4)H^{VPRS} 将几乎一致性规则与一致性规则等同对待,起到了保护几乎一致性规则的作用,从而有效处理了噪声对数据一致性的影响。

表 1　$Q_i\to d$ 的分布

$B_i\cap D_j$	B_1	B_2	B_3	B_4	$\sum$	$T_i\cap D_j$	T_1	T_2	T_3	T_4	$\sum$
D_1	25	25	25	25	100	D_1	46	4	25	25	100
D_2	25	25	25	25	100	D_2	4	46	25	25	100
$\sum$	50	50	50	50	200	$\sum$	50	50	50	50	200
$X_i\cap D_j$	X_1	X_2	X_3	X_4	$\sum$	$Y_i\cap D_j$	Y_1	Y_2	Y_3	Y_4	$\sum$
D_1	46	4	50	0	100	D_1	50	0	50	0	100
D_2	4	46	0	50	100	D_2	0	50	0	50	100
$\sum$	50	50	50	50	200	$\sum$	50	50	50	50	200
$Z_i\cap D_j$	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	$\sum$		
D_1	23	2	23	2	25	0	25	0	100		
D_2	2	23	2	23	0	25	0	25	100		
$\sum$	25	25	25	25	25	25	25	25	100		

表 2　对 $Q_i\to d$ 的 5 种不确定性量度计算结果

$Q_i\to d$	$Q_1\to d$	$Q_2\to d$	$Q_3\to d$	$Q_4\to d$	$Q_5\to d$
γ_0	0	0	0.5	1	0.5
H^{det}	7.6439	7.6439	4.8219	2	5.3219
H^{VPRS}	7.6439	4.8219	2	2	3

4　结束语

从信息系统中发现不确定性的知识是粗糙集理论的独特之处,同时也是粗糙集在数据挖掘应用中最具特色之处。文中讨论了几种粗糙决策规则集量度的方法,给出了变精度粗糙集意义下基于熵的不确定性量度函数,并用一个例子比较了 γ_0、H^{det} 和 H^{VPRS} 的优劣。结果说明,基于信息熵的不确定性量度函数不但能兼顾规则的一致性和随机性,而且还有效地处理了噪声对数据一致性的影响,对几乎一致性规则起到了保护作用。在数据挖掘中,不确定性量度也可被用作进行规则挖掘和决策模型选择的依据。

(下转第 57 页)

表 1　仿真参数

参数名	参数值
信道比特速率	1Mbps
时隙 δ 长度	20μs
SIFS	10μs
DIFS	50μs
PIFS	30μs
EIFS	212μs
CW_{min}	31
CW_{max}	1023
PHY 头大小	192b
MAC 头大小	144b
RTS 大小	160b
CTS 大小	112b
ACK 大小	112b
DLP－Duration	12ms
DCF－Duration	8ms

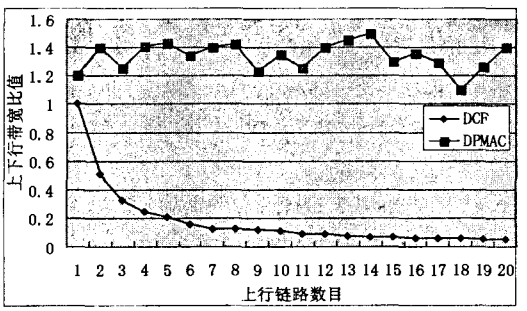

图 3　DCF 和 DPMAC 仿真结果

4　结束语

建立了三维 Markov 链模型,对非饱和状态下 802.11 MAC 协议造成的上下行带宽公平性进行了研究,由此提出基于下行链路优先的 DPMAC 协议,通过仿真证实 DPMAC 协议能很好地实现上下行链路公平性,达到满意的效果。

参考文献:

[1] Heusse M, Rousseau F. Performance Anomaly of 802.11b [C]// Proceedings of IEEE INFOCOM'03. San Francisco, California:[s.n.],2003:231－238.

[2] Pang Q X, Liew S C, Lee Y B. A TCP－like Adaptive Contention Window Scheme for WLAN[C]// Proc. of IEEE International Conference on Communications. Paris, France: [s.n.], 2004.

[3] Deng J, Varshney P K, Haas Z J. A New Backoff Algorithm for the IEEE 802.11 Distributed Coordination Function[C] // Proc. of Communication Networks and Distributed Systems Modeling and Simulation. San Diego, CA, USA:[s.n.], 2004.

[4] Cali F, Conti M, Gregori E. IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism[J]. IEEE JSAC, 2000,18(9):1774－1786.

[5] 冯　辉,王　挺,胡　波.一种 WLAN 中优化上下行公平性的 MAC 机制[J].复旦学报:自然科学版,2006,45(1):67－71.

[6] Bottigliengo M, Casetti C, Chiasserini C F, et al. Smart traffic scheduling in 802.11 WLANs with access point[C]// CERCOM － Dipartimento di Elettronica Politecnico di Torino. Torino, Italy:[s.n.], 2003.

[7] Nandiraju N S P, Gossain H, Cavalcanti D, et al. Achieving Fairness in Wireless LANs by Enhanced IEEE 802.11 DCF [C] // Wireless and Mobile Computing, Networking and Communications, 2006(WiMob'2006). IEEE International Conference.[s.l.]:[s.n.],2006:132－139.

[8] Kim Sung Won, Kim Byung－Seo. Fang Yuguang. Downlink and uplink resource allocation in IEEE 802.11 wireless LANs [J]. IEEE Transaction and Vehicular Technology, 2005, 54 (1):320－327.

[9] Qiao Daji, Shin K G. Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 WLAN under the DCF[C]// The 10th IEEE International Workshop on Quality of Service. USA: IEEE, 2002.

[10] Yoo Joon, Luo Haiyun, Kim Chong－kwon. Opportunistic Joint Uplink/Downlink Scheduling in WLANs[R]. US: University of Illinois at Urbana－Champaigny, Korea: Seoul National University, 2006.

[11] 李　波,李建东,方　勇.非饱和状态下 IEEE 802.11 DCF 的性能分析[J].西安电子科技大学学报,2007,34(1):76－81.

[12] 刘乃安.无线局域网(WLAN)——原理、技术与应用[M].西安:西安电子科技大学出版社,2004:306－307.

(上接第 53 页)

参考文献:

[1] Pawlak Z, Grzymala－Busse J, Slowin ski R, et al. Rough sets[J]. Communications of the ACM, 1995, 38(11): 88－95.

[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data[M]. Dordrecht: Kluwer Academic Publishers, 1991.

[3] Düntsch I, Gediga G. Uncertainty measures of rough set prediction[J]. Artificial Intelligence, 1998, 106: 109－137.

[4] Ziarko W. Variable precision rough set model[J]. Journal of Computer and System Sciences, 1993, 46: 39－59.

[5] 胡寿松, 何亚群. 粗糙决策理论与应用[M].北京:北京航空航天大学出版社, 2005.

[6] 陶　志. 基于粗糙集理论的数据挖掘方法及其在电力营销决策支持系统中的应用[D].沈阳:东北大学,2004.