

覆盖粗糙 Vague 集的不确定性度量研究

徐久成 张倩倩

(河南师范大学计算机与信息技术学院 新乡 453007)

摘 要 在覆盖粗糙集理论研究的基础上,将粗糙集理论与 Vague 集理论相融合,研究基于覆盖的粗糙 Vague 集模型及其相关性质。为了更好地度量该模型的不确定性,首先定义了知识含量测度,用于度量覆盖对论域中对象分类能力产生的不确定性,此种度量方法体现了分类知识的本质特征;同时结合边界域大小的度量方法,定义了 C-粗糙度的概念用于度量覆盖粗糙 Vague 集的不确定性。实例分析结果表明了这种度量方法的简洁高效性。

关键词 Vague 集,粗糙 Vague 集,知识含量,不确定性

中图法分类号 TP18 文献标识码 A

Research on Uncertainty Measurement for Covering Rough-Vague Sets

XU Jiu-cheng ZHANG Qian-qian

(College of Computer & Information Technology, Henan Normal University, Xinxiang 453007, China)

Abstract On the researching of covering rough sets theory, Rough-Vague sets model and its properties based on cover were researched by integrating rough sets theory and vague sets theory. In order to measured this model's uncertainty perfectly, firstly, it measured the uncertainty of the classification capability of cover by defining knowledge capacity measure. Such method reflects the essential characteristic of knowledge classification. And then it defined C-roughness to measure the model's uncertainty effectively by combining knowledge capacity and roughness (the measure of border field). The concisely and validity for this measurement were proved by analyzing example.

Keywords Vague sets, Rough-Vague sets, Knowledge capacity, Uncertainty

1 引言

粗糙集理论^[1]和 Vague 集理论^[2]都是用于处理含糊的、不确定性问题的理论工具,但二者的出发点和着眼点不同,各具优势,在各自应用领域发挥着重要作用。Vague 集理论虽然和 Fuzzy 集一样都可用于边界不确定信息的处理^[3],但与 Fuzzy 集相比较能更好地表达模糊信息。由于实际应用的需要,有学者将粗糙集模型推广到模糊领域,使得粗糙集和 Vague 集自然融合,提出了粗糙 Vague 集模型^[4-6],推广和发展了粗糙集理论,为进一步研究不确定理论提供了支撑。

粗糙集模型中,等价关系是一个限制条件,但是在实际应用中,等价关系的构造是很困难的,这就限制了粗糙集的应用。因此,我们将严格的等价关系推广到一般的二元关系,就会构成论域的覆盖^[7]。覆盖是划分的推广,其所形成的对象子集之间存在相交的情况,比划分更具普遍性,将是极具应用价值的研究方向。本文就是在覆盖空间关系下进一步研究粗糙 Vague 集模型,拓广其应用范围。那么,如何度量覆盖粗糙 Vague 集的不确定性呢?文献[8]指出,粗糙集的不确定性主要由两个原因引起,一个来自于给定近似空间的粗糙集的边界,另一个来自于论域上的二元关系对论域中对象的分

类能力,也就是知识的粒度大小所引起的不确定性。因而对于覆盖粗糙 Vague 集的不确定性度量也必然要同时考虑这两方面影响因素。首先对于知识粒度所引起的不确定性,多是采用信息论中熵来度量的^[9,10],度量结果具有相对性,并且实际运算中比较复杂,因此本文提出了一种知识含量测度,从而有效地度量了覆盖近似空间中知识的不确定性,其运算过程简便,并且指出知识含量会随着覆盖的变细而增大。同时结合边界域大小的度量方法定义了 C-粗糙度的概念,用来度量覆盖粗糙 Vague 集的不确定性。

2 Vague 集相关理论

定义 1^[7] 设 U 为非空有限论域, $C = \{X | X \subseteq U\}$ 是 U 的子集族,如果:对 $\forall X \in C, X \neq \emptyset$, 而且 $\bigcup C = U$, 则称 C 是 U 的一个覆盖,并称 (U, C) 为覆盖近似空间。

显然,上述覆盖的定义中不存在子集之间相交为空的条件,因此,覆盖是划分的推广,论域 U 的划分是 U 的一个覆盖。

定义 2^[7] 设 (U, C) 是一个覆盖近似空间, $x \in U$, 称 $Md(x) = \{K \in C | x \in K \wedge (\forall S \in C \wedge x \in S \wedge S \subseteq K \Rightarrow K = S)\}$ 为 x 的最小描述。

到稿日期:2009-11-12 返修日期:2010-01-26 本文受国家自然科学基金项目(60873104),河南省高校新世纪优秀人才支持计划(2006HANCET-19),河南省教育厅自然科学基金项目(2008B520019)资助。

徐久成(1963—),男,博士,教授,主要研究方向为粗糙集理论、粒计算、数据挖掘等, E-mail: xjc@henannu.edu.cn; 张倩倩(1982—),女,硕士生,主要研究方向为粗糙集、Vague 集理论与粒计算。

定义 3^[11] 设论域 $U = \{u_1, u_2, \dots, u_n\}$, 其中元素 $u_i (i = 1, 2, \dots, n)$ 是所讨论的对象。 U 上一 Vague 集 A 由真隶属度函数 t_A 和假隶属度函数 f_A 所描述: $t_A: U \rightarrow [0, 1], f_A: U \rightarrow [0, 1]$ 。其中 $t_A(u_i)$ 是由支持 u_i 的证据所导出的肯定隶属度的下界, $f_A(u_i)$ 则是由反对 u_i 的证据所导出的否定隶属度的下界, 且 $t_A(u_i) + f_A(u_i) \leq 1$ 。元素 u_i 在 Vague 集 A 中的隶属度被区间 $[0, 1]$ 的一个子区间 $[t_A(u_i), 1 - f_A(u_i)]$ 所界定, 称该区间为 u_i 在 A 中的 Vague 值。

定义 4^[11] 一个 Vague 集 A 的补集 A^c 定义为: $t_{A^c}(x) = f_A(x), 1 - f_{A^c}(x) = 1 - t_A(x)$ 。

性质^[11] Vague 集 A 为 B 所包含, 即 $A \subseteq B$, 当且仅当 $t_A(x) \leq t_B(x)$ 且 $1 - f_A(x) \leq 1 - f_B(x)$ 。

定义 5^[11] 两个 Vague 集 A 和 B 是相等的, 即 $A = B$, 当且仅当 $A \subseteq B$ 和 $B \subseteq A$, 即 $t_A(x) = t_B(x); 1 - f_A(x) = 1 - f_B(x)$ 。

3 基于覆盖的粗糙 Vague 集模型及其性质

定义 6 设 (U, C) 是一个覆盖近似空间, 其中 U 为非空有限论域, C 为 U 上的一个覆盖, $\forall x \in U, U$ 上的一个 Vague 集 A 关于覆盖近似空间 (U, C) 的下近似与上近似分别为 $\underline{C}(A), \bar{C}(A)$, 则序对 $CV = (\underline{C}(A), \bar{C}(A))$ 为基于覆盖的粗糙 Vague 集, 其真、假隶属度函数分别定义为:

$$\begin{aligned} t_{\underline{C}(A)}(x) &= \inf\{t_A(y) \mid y \in \bigcup Md(x)\} \\ f_{\bar{C}(A)}(x) &= \sup\{f_A(y) \mid y \in \bigcup Md(x)\} \\ t_{\bar{C}(A)}(x) &= \sup\{t_A(y) \mid y \in \bigcup Md(x)\} \\ f_{\underline{C}(A)}(x) &= \inf\{f_A(y) \mid y \in \bigcup Md(x)\} \end{aligned}$$

在定义 6 中, 若对 $\forall x \in U$ 有 $\underline{C}(A)(x) = \bar{C}(A)(x)$, 则称 Vague 集 A 关于覆盖近似空间 (U, C) 是可定义的, 否则称 Vague 集 A 关于覆盖近似空间 (U, C) 是粗糙的, 此时也称 A 为粗糙 Vague 集。这里我们称 $\underline{C}, \bar{C}$ 分别为覆盖粗糙 Vague 集的下近似算子与上近似算子。

从定义 6 可以看出, 在覆盖空间 (U, C) 上定义的粗糙 Vague 集表示的是在覆盖知识 C 上, 用论域上一对可定义的下近似和上近似子集来逼近所讨论的模糊对象集合, 从而可以发现模糊信息系统中一些隐藏的知识。

在此规定基于覆盖的粗糙 Vague 集 $CV = \emptyset$, 当且仅当 $\underline{C}(A) = \emptyset$ 且 $\bar{C}(A) = \emptyset$ 。

设 (U, C) 是一个覆盖近似空间, A, B 分别为论域 U 上的 Vague 集, 则覆盖粗糙 Vague 集下近似算子 $\underline{C}$ 与上近似算子 $\bar{C}$ 具有如下性质:

性质 2 $\underline{C}(\emptyset) = \bar{C}(\emptyset) = \emptyset, \underline{C}(U) = \bar{C}(U) = U$ 。

性质 3 $\underline{C}(A) \subseteq A \subseteq \bar{C}(A)$ 。

性质 4(可加性 1) $\underline{C}(A \cap B) = \underline{C}(A) \cap \underline{C}(B)$;

性质 5(可加性 2) $\bar{C}(A \cup B) = \bar{C}(A) \cup \bar{C}(B)$ 。

性质 6(单调性) $A \subseteq B \Rightarrow \underline{C}(A) \subseteq \underline{C}(B), \bar{C}(A) \subseteq \bar{C}(B)$ 。

性质 7 $\underline{C}(A^c) = (\bar{C}(A))^c; \bar{C}(A^c) = (\underline{C}(A))^c$ 。

性质 8(幂等性) $\underline{C}(A) = \underline{C}(\underline{C}(A)); \bar{C}(A) = \bar{C}(\bar{C}(A))$ 。

证明: 由定义 6 知, $t_{\underline{C}(A)}(x) = \inf\{t_{\underline{C}(A)}(y) \mid y \in \bigcup Md(x)\} = \inf\{\inf\{t_A(y) \mid y \in \bigcup Md(x)\} \mid y \in \bigcup Md(x)\} = t_{\underline{C}(A)}(x)$;

$f_{\bar{C}(A)}(x) = \sup\{f_{\bar{C}(A)}(y) \mid y \in \bigcup Md(x)\} = \sup\{\sup\{f_A(y) \mid y \in \bigcup Md(x)\} \mid y \in \bigcup Md(x)\} = \sup\{f_A(y) \mid y \in \bigcup Md(x)\} = f_{\bar{C}(A)}(x)$,

所以 $\underline{C}(A) = \underline{C}(\underline{C}(A))$ 。同理有 $\bar{C}(A) = \bar{C}(\bar{C}(A))$ 。

性质 9 当覆盖 C 是最细的粒度关系, 即 $C = \{\{x_1\}, \{x_2\}, \{x_3\}, \dots, \{x_n\}\}$ 时, 有 $\underline{C}(A) = A = \bar{C}(A)$ 。

性质 9 表明, 当覆盖 C 的粒度最细时, 基于覆盖的粗糙 Vague 集是可定义的。

4 基于知识含量的覆盖粗糙 Vague 集不确定性度量

在粗糙集理论中, 粗糙集的不确定性度量是一项基础性研究工作, 对于粗糙集的应用起到非常重要的作用。基于覆盖的粗糙 Vague 集模型是对经典粗糙集的推广, 因而探讨该模型的不确定性度量方法是很有必要的。

引起覆盖粗糙 Vague 集不确定性的原因之一是覆盖近似空间对论域中对象的分类能力, 在粗糙集理论中多是采用知识的信息熵来度量知识分类的不确定性, 运算过程中使用对数, 计算过程复杂, 且计算结果具有相对性。在此, 为了体现覆盖粒度空间中分类知识的本质特征, 定义知识含量测度来寻求一种更简单的覆盖知识的不确定性度量方法。

文献[12]指出, 论域中的元素由于不可分辨性构成了覆盖粒度空间中的粒子。对于论域中的某个元素, 可能同时属于多个粒子, 这是人们从不同的角度观察事物的结果。对于论域中的两个元素, 可能它们从某个角度观察是不可区分的, 但从其他的角度观察却是可区分的。因此, 对于论域中的某个元素, 在覆盖粒度空间中的不确定性应由各个角度都不可分辨的元素决定, 文中给出了不可分辨集的定义。

定义 7 设 (U, C) 为一覆盖近似空间, x 为 U 中的一个对象, 则在覆盖近似空间下 x 的描述集为 $Des_C(x) = \{K \mid K \in C \wedge x \in K\}$ 。

定义 8 设 (U, C) 为一覆盖近似空间, x 为论域 U 中的一个对象, 则在覆盖近似空间下 x 的不可分辨集定义为: $Ind_C(x) = \bigcap_{K \in Des_C(x)} K$ 。

下面, 我们来定义覆盖近似空间的知识含量。

定义 9 设 (U, C) 为一覆盖空间, $U = \{x_1, x_2, \dots, x_n\}$ 为对象的论域, $C = \{K_1, K_2, \dots, K_m\}$ 是 U 上的一个覆盖, 定义 $I(C) = \begin{cases} 1, & |K_i| = 1, i = 1, 2, \dots, m \\ 1 - \sum_{x_i \in U} |Ind_C(x_i)| / |U|^2, & \text{其它} \end{cases}$

为覆盖 C 的知识含量测度, 其中 $|\cdot|$ 表示集合的基数。

从定义 9 可以看出, 当覆盖 C 的粒度最细, 即 $|K_i| = 1$ 时, 覆盖的知识含量最大为 1; 当覆盖为最粗的论域关系时, 知识含量最小为 0。

由于划分是覆盖的一个特例, 因此 Pawlak 划分近似空间是覆盖近似空间的特例。所以, 当覆盖近似空间退化为经典的划分近似空间时, 定义 9 中所定义的覆盖近似空间的知识含量就应该退化为划分近似空间下的知识含量。显然, 当 C 为论域上的划分时, 定义 9 给出的覆盖知识含量 $I(C)$ 与文献[13, 14]中所定义的基于论域上等价关系划分近似空间下的知识含量一致。

性质 10 设 (U, C) 为一覆盖近似空间, 对任一覆盖 C , 有 $0 \leq I(C) \leq 1$ 。特别地, 当 $C = \{U\}$ 时, $I(C) = 0$; 当 $C = \{\{x_1\}, \{x_2\}, \{x_3\}, \dots, \{x_n\}\}$ 时, $I(C) = 1$ 。

由定义 9 可知性质 10 显然成立, 证明略。

定义 10^[15] 设 U 为非空有限论域, C_1 和 C_2 为 U 上的

两个覆盖,如果 $\forall K_i \in C_1$, 存在 $K_j \in C_2$, 使得 $K_i \subseteq K_j$, 且 $\forall K_j \in C_2$, 存在 $K_i \in C_1$, 使得 $K_i \subseteq K_j$, 则称覆盖 C_1 较覆盖 C_2 细, 记为 $C_1 \leq C_2$ 。

性质 11 设 $(U, C_1), (U, C_2)$ 为两个覆盖近似空间, 若 $C_1 < C_2$, 则 $I(C_1) > I(C_2)$ 。

证明: 如果 $C_1 < C_2$, 则有 $\forall K_i \in C_1$, 存在 $K_j \in C_2$, 使得 $K_i \subset K_j$, 且 $\forall K_j \in C_2$, 存在 $K_i \in C_1$, 使得 $K_i \subset K_j$, 则 $\forall x_i \in U$, 有 $Desc_1(x_i) \subset Desc_2(x_i)$, 其中 $Desc_1(x) = \{K_i | K_i \in C_1 \wedge x \in K_i\}$, $Desc_2(x) = \{K_j | K_j \in C_2 \wedge x \in K_j\}$ 。则由定义 8 有 $Ind_{C_1}(x_i) \subset Ind_{C_2}(x_i)$, 即 $|Ind_{C_1}(x_i)| < |Ind_{C_2}(x_i)|$, $1 - \sum_{x_i \in U} |Ind_{C_1}(x_i)| / |U|^2 > 1 - \sum_{x_i \in U} |Ind_{C_2}(x_i)| / |U|^2$, 由定义 9 得 $I(C_1) > I(C_2)$ 。

特别地, 对于 U 上的两个覆盖 C_1, C_2 , 若 $\forall x \in U$, 有 $Ind_{C_1}(x) = Ind_{C_2}(x)$, 则 $I(C_1) = I(C_2)$ 。

知识含量主要用来描述覆盖粒度空间所包含的确定性信息的多少, 性质 11 表明知识含量随着覆盖粒度的变细而变大, 即覆盖知识粒度越细, 其包含的知识量越大, 对论域中对象的分辨能力越强, 因而可用覆盖的知识含量测度来对不同的覆盖所引起的知识不确定性进行数值特征刻画。

还有另一个原因会引起覆盖粗糙 Vague 集的不确定性, 就是粗糙 Vague 集边界区域的大小, 这是用一对上、下近似来刻画的。下面给出基于覆盖的粗糙 Vague 集 CV 的精度和粗糙度的定义。

定义 11 设 $CV = (\underline{C}(A), \bar{C}(A))$ 是给定论域 U 上的覆盖粗糙 Vague 集, $0 < \beta \leq \alpha \leq 1$, 则定义 Vague 集 A 的 α 下近似 $\underline{C}(A)_\alpha = \{y | y \in \bigcup Md(x_i), t_{C(A)}(y) \geq \alpha, f_{C(A)}(y) < \alpha, i = 1, 2, \dots, n\}$, Vague 集 A 的 β 上近似 $\bar{C}(A)_\beta = \{y | y \in \bigcup Md(x_i), t_{C(A)}(y) \geq \beta, f_{C(A)}(y) < \beta, i = 1, 2, \dots, n\}$ 。

定义 12 基于覆盖的粗糙 Vague 集的精度 $\gamma_{CV}^{\alpha, \beta}$ 及粗糙度 $\rho_{CV}^{\alpha, \beta}$ 分别定义为:

$$\gamma_{CV}^{\alpha, \beta} = |\underline{C}(A)_\alpha| / |\bar{C}(A)_\beta|, \rho_{CV}^{\alpha, \beta} = 1 - \gamma_{CV}^{\alpha, \beta}$$

式中, $0 < \beta \leq \alpha \leq 1, \bar{C}(A)_\beta \neq \emptyset$ 。

在定义 12 中, 若 $\gamma_{CV}^{\alpha, \beta} = 1$, 则 CV 是可定义的, 即获得完全确定的信息。

基于以上两方面原因的考虑, 定义 13 给出了基于知识含量的覆盖粗糙 Vague 集不确定性度量方法。

定义 13 设 (U, C) 为一个覆盖近似空间, $CV = (\underline{C}(A), \bar{C}(A))$ 为 Vague 集 A 关于论域 U 上的一个覆盖粗糙 Vague 集, $0 < \beta \leq \alpha \leq 1$, 定义基于知识含量的覆盖粗糙 Vague 集的 C -粗糙度为:

$$\rho^{C, \beta}(A, C) = \rho_{CV}^{\alpha, \beta}(1 - I(C))$$

从定义 13 可以看出, 覆盖粗糙 Vague 集的 C -粗糙度, 不仅和它自身的粗糙度大小有关, 还和覆盖的知识含量有关。这一定义能更准确地反映出基于覆盖的粗糙 Vague 集的不确定性程度, 运行过程简便有效。

实例 设有覆盖近似空间 (U, C) , 论域 $U = \{x_1, x_2, \dots, x_7\}$, 已知 U 上的两个覆盖: $C_1 = \{\{x_1, x_2\}, \{x_3, x_6\}, \{x_3, x_7\}, \{x_4, x_5\}, \{x_6\}, \{x_7\}\}$, $C_2 = \{\{x_1, x_2\}, \{x_3, x_6, x_7\}, \{x_3, x_7\}, \{x_4, x_5, x_6\}\}$, 所讨论的 Vague 集对象 $A = \{[0.7, 0.8]/x_1, [0.5, 0.7]/x_2, [0.8, 0.9]/x_3, [0.3, 0.4]/x_4, [0.4, 0.5]/x_5, [0.5, 0.6]/x_6, [0.7, 0.8]/x_7\}$ 。

由定义 10 知 $C_1 < C_2$, 计算覆盖的知识含量有 $I(C_1) = 40/49, I(C_2) = 34/49$, 可见粒度越细的覆盖, 包含的知识量越多。

由定义 6 可知, 序对 $CV_1 = (\underline{C}_1(A), \bar{C}_1(A))$ 为覆盖知识空间 C_1 下的粗糙 Vague 集, 其中:

$$\underline{C}_1(A) = \{[0.7, 0.8]/x_1, [0.5, 0.7]/x_2, [0.5, 0.6]/x_3, [0.3, 0.4]/x_4, [0.3, 0.4]/x_5, [0.5, 0.6]/x_6, [0.7, 0.8]/x_7\}$$

$$\bar{C}_1(A) = \{[0.7, 0.8]/x_1, [0.5, 0.7]/x_2, [0.8, 0.9]/x_3, [0.4, 0.5]/x_4, [0.4, 0.5]/x_5, [0.5, 0.6]/x_6, [0.7, 0.8]/x_7\}$$

为说明定义 13 中不确定性度量方法的有效性, 这里取 $\alpha = 0.7, \beta = 0.4$, 则 $\rho_{CV_1}^{\alpha, \beta} = 1 - |\{x_1, x_7\}| / |\{x_1, x_2, x_3, x_6, x_7\}| = 3/5$; 同样的, 序对 $CV_2 = (\underline{C}_2(A), \bar{C}_2(A))$ 为基于覆盖 C_2 的粗糙 Vague 集, 其中:

$$\underline{C}_2(A) = \{[0.5, 0.7]/x_1, [0.5, 0.7]/x_2, [0.7, 0.8]/x_3, [0.3, 0.4]/x_4, [0.3, 0.4]/x_5, [0.3, 0.4]/x_6, [0.7, 0.8]/x_7\}$$

$$\bar{C}_2(A) = \{[0.7, 0.8]/x_1, [0.7, 0.8]/x_2, [0.8, 0.9]/x_3, [0.5, 0.6]/x_4, [0.5, 0.6]/x_5, [0.8, 0.9]/x_6, [0.8, 0.9]/x_7\}$$

则有 $\rho_{CV_2}^{\alpha, \beta} = 1 - |\{x_3, x_7\}| / |\{x_1, x_2, x_3, x_6, x_7\}| = 3/5$ 。

可以看出, 两个粒度粗细不同的覆盖粗糙 Vague 集却具有了相同的粗糙度, 此时无法准确判断 CV_1 和 CV_2 的不确定性, 这是因为我们忽略了分类知识对模型不确定性造成的影响, 因此使用定义 13 给出的覆盖粗糙 Vague 集的 C -粗糙度的定义, 则有: $\rho^{C, \beta}(A, C_1) = \rho_{CV_1}^{\alpha, \beta}(1 - I(C_1)) = \frac{3}{5} \times \frac{9}{49} = 0.11$; $\rho^{C, \beta}(A, C_2) = \rho_{CV_2}^{\alpha, \beta}(1 - I(C_2)) = \frac{3}{5} \times \frac{15}{49} = 0.184$, 显然, $\rho^{C, \beta}(A, C_1) < \rho^{C, \beta}(A, C_2)$ 。

实例分析结果表明, 仅用粗糙度并不能正确反映覆盖粗糙 Vague 集的不确定性, 必须要考虑到覆盖分类知识的粒度粗细对覆盖粗糙 Vague 集的不确定性造成的影响, 这里知识含量测度正是体现了知识分类的本质特征, 因而使用基于知识含量的 C -粗糙度能更好地度量其不确定性, 并且运算起来更简便。

结束语 覆盖广义粗糙集是数据挖掘领域最具应用前景的一种粗糙集推广模型, 在此进一步通过将粗糙集和 Vague 集两种不确定性理论相结合, 研究覆盖知识空间下粗糙集模型的推广。文中对基于覆盖的粗糙 Vague 集不确定性度量的研究, 有助于发现模糊信息系统中隐藏的知识, 对处理 Vague 信息系统的属性约简、知识获取起到推动作用, 同时也是我们接下来要研究的内容。将粗糙集和 Vague 集理论相结合, 优势互补, 对进一步探索粗糙集和 Vague 集理论具有一定意义。

参考文献

- [1] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 314-356
- [2] Gau W L, Buehrer D J. Vague sets [J]. IEEE Trans Syst Man Vybern, 1993, 23(2): 610-614

(下转第 282 页)

表 2 是广域网智能存储系统集合带宽随系统规模扩大而增长的情况。测试过程中将存储节点数由 1 个增加到 6 个,并模拟 100 个并发用户的访问,测试不同系统规模情况下集合带宽的变化情况。结果表明当存储节点的数目不断增加,系统的集合带宽拟线性增长,这一特点在图 7 中表现得尤为明显,这充分体现了广域网智能存储系统扩容与增速同步的思想。另一方面,广域网智能存储系统中实现了命令与数据的分流,避免了大量的数据在存储节点和元数据服务器之间往返拷贝,将高性能的存储设备和网络带宽充分提供给网络用户,而不被传统文件服务器的瓶颈所限制。

表 2 改进系统集合带宽随存储节点数增加而增长情况

并发用户数	存储节点数目	集合带宽(MB/s)	
		上传	下载
100	1	14.41	109.98
	2	33.76	220.61
	3	73.52	331.24
	4	86.74	420.39
	5	135.15	530.36
	6	153.06	624.67

从表 2 的数据中我们还可以看到,当多用户并发上传数据时,网络已经不是影响传输性能的主要因素,集合带宽受制于存储节点的 I/O 能力;当多用户并发下载数据时,系统的集合带宽比较稳定,基本能达到网络传输的峰值。因而,如果能进一步提高存储节点的写性能,将会使整个存储系统数据上传的能力得到很大提高。

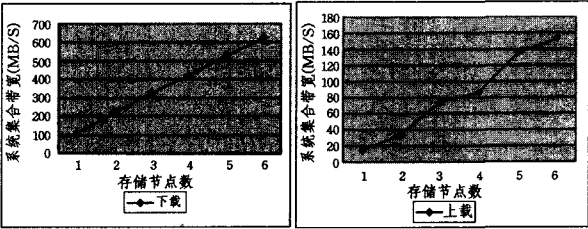


图 7 集合带宽随系统规模扩大拟线性增长示意图

结束语 广域网智能存储系统采用了多层次、可扩展的分布式存储模式,从改进系统结构着手整体地提高网络存储

系统性能,提出基于异构多通道存储设备的、命令与数据分流的网络存储系统,该系统将存储节点通过网络接口直接接入下一代互联网,建立存储节点与用户间的直接数据传输,实现扩容与增速同步,保证服务器与存储节点、存储节点与存储节点间高效的控制和管理。其存储管理遵循 SMI-S 的管理规范,可以在广域网存储系统中提供标准化的通信方式。创新性地设计了广域网存储中间件,以处理复杂网络环境下的数据传输问题,兼顾了广域网数据传输的高速和安全。

参 考 文 献

[1] Seung Woo Son, Chen Guangyu, Kandemir M, et al. Energy Savings through Embedded Processing on Disk System[C]// Proceedings of the 2006 Asia and South Pacific Design Automation Conference, Yokohama, Japan, 2006:128-133

[2] 刘群. 基于可扩展对象的海量存储系统研究[D]. 武汉:华中科技大学图书馆, 2006

[3] Deng Ze, Feng Dan, Shi Zhan, et al. Scalability Support for SMI-S with Chord[C]// Proceeding of the 10th IEEE International Conference on High Performance Computing and Communications, Dalian, China, 2008:219-225

[4] 王娟, 冯丹, 王芳, 等. 一种元数据服务器集群的负载均衡算法[J]. 小型微型计算机系统, 2009, 30(4):757-760

[5] Brandt S A, Miller E L, Long D D E, et al. Efficient Metadata Management in Large Distributed Storage Systems[C]// Proceedings of the 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies, San Diego, California, USA, 2003:290-298

[6] Kher V, Seppanen E, Leach C, et al. SGFS: Secure, Efficient and Policy-based Global File Sharing[C]// Proceedings of the 23rd IEEE/14th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2006), College Park, MD, 2006

[7] Kaminsky M, Savvides G, Mazieres D, et al. Decentralized user authentication in a global file system[C]// Proceedings of the 19th ACM Symposium on Operating Systems Principles, Bolton Landing, New York, 2003:60-73

(上接第 227 页)

[3] 徐久成, 安秋生, 王国胤, 等. 边界不确定信息的处理—Fuzzy 集和 Vague 集[J]. 计算机工程与应用, 2002, 16:24-26, 82

[4] 闫德勤, 迟忠先. 粗糙集与 Vague 集[J]. 计算机科学, 2004, 31(8):133-135

[5] 刘金良, 闫瑞霞, 姚炳学. 粗糙 Vague 集的不确定性度量[J]. 系统工程与电子技术, 2008, 30(1):104-107

[6] 朱六兵. 粗糙集与 Vague 集的理论及应用研究[D]. 成都:西南交通大学, 2006

[7] Zhu W, Wang Feiyue. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences, 2003, 152:217-230

[8] 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

[9] 徐菲菲, 苗夺谦, 李道国, 等. 基于覆盖的粗糙模糊集的粗糙嫡

[J]. 计算机科学, 2006, 33(10):179-181

[10] 张倩倩, 徐久成, 胡玉文. 基于覆盖的粗糙 Vague 集模型研究[J]. 广西大学学报, 2009, 35(5):653-657, 662

[11] 李凡, 徐章艳, 饶勇. Vague 集[J]. 计算机科学, 2000, 27(9):12-14, 28

[12] 胡军, 王国胤. 覆盖粒度空间的层次模型[J]. 南京大学学报, 2008, 44(5):551-558

[13] Liang Jiye, Li Deyu. Information measure of roughness of knowledge and significance of attribute in rough set theory[J]. Journal of Engineering Mathematics, 2000, 17(supp):106-108

[14] 刘纪芹, 史开泉. 基于知识含量的粗糙集不确定性度量[J]. 计算机科学, 2007, 34(7):171-173

[15] Zhang Q H, Wang G Y, Hu J, et al. Approximation partition spaces of covering space[C]//IEEE International Conference on Granular Computing, Silicon Valley, 2007:199-204