

Vague 集中的分解定理与相似度量

闫德勤¹ 迟忠先²

(辽宁师范大学计算机系 大连116029)¹ (大连理工大学计算机系 大连116026)²

Decomposition Theorem and Measures of Similarity in Vague Sets

YAN De-Qin¹ CHI Zhong-Xian²

(Department of Computer Science, Liaoning Normal University, Dalian 116029)¹

(Department of Computer Science, Dalian University of Technology, Dalian 116026)²

Abstract Based on fuzzy set theory this paper gives decomposition theorem in Vague sets, and by discussing the measures of similarity of Vague sets this paper presents a new kind of the measure. The results can be used for further research and applications of Vague sets.

Keywords Vague sets, Fuzzy sets, Decomposition theorem, Measures of similarity

1 引言

目前,在计算机科学及应用的多种领域中,模糊集起着重要的作用。因而其理论及应用都得到了迅速的发展。Vague 集^[1]从一定意义上讲在对事物属性的描述上较模糊集提供了更多的选择方式,因而已在学术界和工程技术界引起了广泛关注。事实上,Vague 集正成为当前模糊信息处理中的一个新兴课题。当前相对于应用而言,Vague 集的理论研究方面有待于丰富和发展。本文以模糊集理论为基础对 Vague 集进行了探讨,给出了关于 Vague 集的分解定理,同时也给出了 Vague 集中集合的相似度量公式。这些理论结果为对 Vague 集进行进一步的理论研究提供一定基础。

2 基本概念

定义1^[1] 设 X 是一个对象空间,其中的任意一个元素用 x 表示, X 中的一个 Vague 集 V 用一个真隶属函数 t_v 和一个假隶属函数 f_v 表示。 $t_v(x)$ 是从支持 x 的证据所导出的 x 的隶属度下界, $f_v(x)$ 则是从反对 x 的证据所导出的 x 的否定隶属度下界, $t_v(x)$ 和 $f_v(x)$ 将区间 $[0,1]$ 中的一个实数与 X 中的每一个点联系起来,即

$$t_v: X \rightarrow [0,1]$$

$$f_v: X \rightarrow [0,1]$$

x 关于 V 的隶属度 $V(x)$ 表示为:

$$[t_v(x), 1 - f_v(x)].$$

其中, $t_v(x) + f_v(x) \leq 1$ 。

当 X 为连续空间时,一个 Vague 集 V 表示为:

$$V = \int_X [t_v(x), 1 - f_v(x)] / x, x \in X.$$

当 X 为离散空间时,一个 Vague 集 V 表示为:

$$V = \sum_{i=1}^n [t_v(x_i), 1 - f_v(x_i)] / x_i, x_i \in X.$$

例1 若 $X = \{1, 2, \dots, 10\}$,则表示数值为“小”的一个 Vague 集 V 可定义为:

$$V = [1, 1] / 1 + [0.9, 1] / 2 + [0.6, 0.8] / 3 + [0.3, 0.5] / 4 + [0.1, 0.2] / 5.$$

定义2^[5] 设 x 为论域 X 中的任一元素,给定映射:

$$m_A(x): X \rightarrow [0, 1],$$

则 m_A 确定了 X 上的一个模糊集合 A 。 m_A 为 A 的隶属度函数, $m_A(x)$ (简称为 $A(x)$)表示 x 关于 A 的隶属度,即 x 属于 A 的程度。

模糊集合的表示与 Vague 集具有类似的方式。实际上 Vague 集是在模糊集的基础上产生的。

例2 若 $X = \{1, 2, \dots, 10\}$,则表示数值为“小”的一个模糊集 F 可定义为:

$$F = 1/1 + 0.9/2 + 0.6/3 + 0.3/4 + 0.1/5.$$

在普通集合的定义下,元素的属性非“此”即“彼”。模糊集^[5]首先突破了这一限制,使得对事物的描述多样化。从这一点上看,Vague 集是对模糊集的发展。然而,自 Vague 集产生以来,其理论的发展远不如模糊集。鉴于 Vague 集与模糊集的特殊关系,我们以模糊集理论为基础给出关于 Vague 集的理论探讨。

3 Vague 集分解定理与相似度量

定义3 设 V 为一 Vague 集, λ 为一二元表示: $(\lambda, \lambda_f), 0 \leq \lambda \leq 1, \lambda \leq \lambda_f \leq 1$ 。称 $V_\lambda = \{x | t_v(x) \geq \lambda, 1 - f_v(x) \geq \lambda_f\}$ 为 Vague 集 V 的截集。

定义4 设 V 为一 Vague 集, λ 为一二元表示: $(\lambda, \lambda_f), 0 \leq \lambda \leq 1, \lambda \leq \lambda_f \leq 1$ 。 λ 与 V 的积 λV 为一 Vague 集:对 X 中的任一元素 x ,

$$t_{\lambda V}(x) = \lambda \wedge t_v(x),$$

$$1 - f_{\lambda V}(x) = \lambda_f \wedge (1 - f_v(x)).$$

定理1(Vague 集中的分解定理) Vague 集中的任一集合 V 可表示为 $V = \bigcup_{\lambda} \lambda V_\lambda$ 。

证明:对 X 中任一元素用 x ,由 V_λ 定义知:

$$V_\lambda(x) = \begin{cases} 1, & (t_v(x) \geq \lambda, 1 - f_v(x) \geq \lambda_f) \\ 0, & \text{其它} \end{cases}$$

从而,

$$\begin{aligned} (\bigcup_{\lambda} \lambda V_\lambda)(x) &= \bigvee_{\lambda} (\lambda \wedge V_\lambda(x)) = \bigvee_{0 \leq \lambda_1 \leq t_v(x), t_v(x) \leq \lambda_1 \leq \lambda_f \leq 1 - f_v(x)} (\lambda_1 \wedge V_{\lambda_1}(x)) \\ &= \bigvee_{0 \leq \lambda_1 \leq t_v(x), t_v(x) \leq \lambda_1 \leq \lambda_f \leq 1 - f_v(x)} \lambda_1 = V(x). \text{证毕。} \end{aligned}$$

闫德勤 博士,副教授,主要研究领域为模式识别、图象处理等。迟忠先 教授,博士生导师,主要研究领域为知识发现、数据仓库、数据挖掘等。

集合的相似度量对其应用有着重要的作用^[2-4],对 Vague 集这里我们给出一种新的相似度量方法。

设 X 中有元素 $x_i (i=1, 2, \dots)$, 对 Vague 集合记:

$$\Sigma_{\text{count}}(A) = \Sigma_i (t_A(x_i) + f_A(x_i))$$

对 Vague 集中任意两集合 A, B 定义其相似度量 $S(A, B)$ 为:

$$S(A, B) =$$

$$1 - \frac{\Sigma_i \max(0, t_A(x_i) - t_B(x_i)) + \max(0, t_B(x_i) - t_A(x_i))}{\Sigma_{\text{count}}(A \cup B)} - \frac{\Sigma_i \max(0, f_A(x_i) - f_B(x_i)) + \max(0, f_B(x_i) - f_A(x_i))}{\Sigma_{\text{count}}(A \cup B)}$$

$$= \frac{\Sigma_{\text{count}}(A \cap B)}{\Sigma_{\text{count}}(A \cup B)}$$

对于相似度量 S 可以验证具有以下性质:

性质1 $0 \leq S(A, B) \leq 1$, 当且仅当 $A=B$ 时, $S(A, B) =$

1.

性质2 $S(A, B) = S(B, A)$ 。

性质3 $S(A \cup B, A \cap B) = S(B, A)$ 。

性质4 $S(A \cup B, C) \leq S(A, C) + S(B, C)$ 。

(上接第59页)

法是一种新的优化技术,它基于生物进化的概念设计这一系列的过程来达到优化的目的。在多媒体开采中,为了应用遗传算法,需要把多媒体开采任务表达为一种搜索问题而发挥遗传算法的优化搜索能力。

3)归纳学习的方法:归纳学习的方法是传统的数据开采方法的重要研究方向。主要包括决策树、粗集(Rough Set)等方法。

4)模糊数学的方法:利用模糊集合理论对实际问题进行模糊评判、模糊决策、模糊模式识别等处理。

5)基于本体论(Ontology)和多主体的方法:在文[9]中,作者提出了一种利用智能主体和本体论理论设计的多媒体信息检索和开采系统,其利用 Ontology 对多媒体数据进行领域分类,同时对用户的查询信息进行规范,较好地完成多媒体开采和检索的任务。

除了以上谈到的方法外,还有象 Bayesian 分类和 Bayesian 学习算法等方法。

以上谈到的这些开采方法基本上是传统的数据开采的方法在多媒体开采中的应用和扩展。如果从多媒体时间、空间、运动特征、视听特性等角度考虑,必然有新的多媒体开采的方法的出现。例如:Web 结构开采是分析页面的结构和 Web 的结构,在此基础上对页面进行分类和聚类从而找到权威的页面。Brin 等人从 Web 的结构出发就提出了不同于传统数据开采方法的 Page-rank 方法^[10]。

总结 本文对新一代多媒体内容管理中的关键技术——多媒体开采的概念和分类、一般的系统框架、开采的过程和方法作了初步的探讨。多媒体开采是多媒体内容分析、多媒体内容管理、数据仓库和数据开采等多领域的一个交叉学科,也是一门新兴的学科。关于多媒体开采的方法、综合多种媒体的多媒体开采、多媒体开采的交互和表现等诸多问题都需要深入的研究。

结论 通过对 Vague 集的探讨本文给出了 Vague 集中的分解定理以及相似性度量公式。近来 Vague 集作为一种新的工具被广泛研究和应用,但处于起步阶段。本文的结果是为发展其理论所做的探讨。

参考文献

- 1 Gau W L, Buehrer D J. Vague sets. IEEE Trans. Systems Man Cybernet, 1993, 23(2): 610~614
- 2 Chen S M. Measures of Similarity Between Vague Sets. Fuzzy Sets and Systems, 1995, 74: 217~223
- 3 Hong D H, Kim C. A Note on Similarity Measures Between Vague Sets and Between Elements. Information Sciences, 1999, 115(1): 83~96
- 4 闫德勤. 新的模糊度量与模糊似然函数. 模式识别与人工智能, 2001, 14(1): 23~26
- 5 郭宗祥, 杨鸿铨. 模糊信息处理基础. 成都: 成都电讯工程学院出版社, 1989. 22~29

参考文献

- 1 Zhang Ji, Hsu W, Lee M L. IMAGE MINING: ISSUES, FRAMEWORKS AND TECHNIQUES. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 2 Uehara Y, et al. A Computer-aided Visual Exploration System for Knowledge Discovery from Images. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 3 Tung A K H, Ng R T, Lakshmanan L V S, Han Jiawei. Geo-spatial Clustering with User-Specified Constraints. In: the 1st Intl. Workshop on MDM, Boston MA, USA, July 2000
- 4 Antonie M L, Zaiane O R. Application of Data Mining Techniques for Medical Image Classification. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 5 Chen Shu-Ching, Shyu Mei-Ling, Zhang Chengcui, Strickrott J. Multimedia Data Mining for Traffic Video Sequences. In: the 2nd Intl. Workshop on Multimedia Data Mining, San Francisco, California, USA, Aug. 2001
- 6 Datcu M, Seidel K. Image information mining: exploration of image content in large archives. IEEE Conference on Aerospace, Vol. 3, 2000
- 7 Osmar R Z, Han Jiawei, et al. MultiMediaMiner: A System Prototype for MultiMedia Data Mining, 1999
- 8 Zhang J, Hsu W, Lee M L. An Information-driven Framework for Image Mining. In: Proc of 12th Intl. Conf. on Database and Expert Systems Applications, Munich, Germany, Sep. 2001
- 9 Simoff S J, Maher M L. Ontology-based multimedia data mining for design information retrieval
- 10 Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. In: Proc of 7th World Wide Web Conf. Brisbane, Austria, 1998
- 11 熊华. 视频结构化技术的研究与实现. [国防科技大学博士论文]. 2001