不确定统计学习理论

哈明虎　王　超
张植明　田大增　◎著

科学出版社
www.sciencep.com

不确定统计学习理论

哈明虎　王　超
张植明　田大增　著

科学出版社
北　京

内 容 简 介

本书较系统地介绍了不确定统计学习理论，除扼要介绍国内外其他学者的研究成果外，主要介绍作者已公开发表和尚未公开发表的系列研究工作．主要内容包括：不确定学习过程的一致性、不确定学习过程收敛速度的界、不确定结构风险最小化原则以及不确定支持向量机.

本书可作为数学、计算机科学与技术和管理科学与工程等专业高年级本科生、研究生的教材或教学参考书，也可供相关领域的科研人员和工程技术人员阅读参考.

图书在版编目(CIP)数据

不确定统计学习理论/哈明虎等著. —北京：科学出版社，2010

ISBN 978-7-03-027787-9

Ⅰ. 不… Ⅱ. 哈… Ⅲ. 统计学 Ⅳ. C8

中国版本图书馆 CIP 数据核字(2010) 第 100120 号

责任编辑：王丽平 房 阳／责任校对：陈玉凤

责任印制：钱玉芬／封面设计：耕者设计工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2010 年 7 月第 一 版 开本：B5(720 × 1000)

2010 年 7 月第一次印刷 印张：12 1/4

印数：1—2 500 字数：233 000

定价：39.00 元

(如有印装质量问题，我社负责调换)

前　言

统计学习理论是 Vapnik 等在 20 世纪 60 年代提出的, 90 年代中期建立的一种利用经验小样本数据进行机器学习的一般理论. 由于其理论体系的完备性和实际应用的广泛性, 统计学习理论备受机器学习及相关领域科研和工程技术人员的青睐, 目前, 它被认为是处理小样本学习问题的最佳理论. 由于统计学习理论是建立在概率测度空间上基于实随机样本的, 故它难以处理客观世界中大量存在的概率测度空间上基于非实随机样本 (模糊、粗糙、复值和集值样本等) 和非概率测度空间 (Sugeno 测度、可能性测度、拟概率、不确定测度以及集值概率空间等) 上基于非实随机样本的机器学习问题. 为了简便起见, 把概率测度和非概率测度 (Sugeno 测度、可能性测度、拟概率、不确定测度以及集值概率等) 统一称为广义不确定测度, 把概率测度空间和非概率测度空间统一称为广义不确定测度空间, 把实随机样本和非实随机样本统一称为广义不确定样本. 本书旨在构建处理广义不确定测度空间上基于广义不确定样本机器学习问题的统计学习理论 (简称为不确定统计学习理论). 不确定统计学习理论是传统统计学习理论的发展和拓广. 本书主要介绍作者已公开发表和尚未公开发表的系列研究工作.

本书的内容安排如下: 第 1 章为绪论, 第 2 章为预备知识, 第 3 章为不确定学习过程的一致性, 第 4 章为不确定学习过程收敛速度的界, 第 5 章为不确定结构风险最小化原则, 第 6 章为不确定支持向量机初步.

诚然, 不确定统计学习理论的研究处于初级阶段, 很多内容尚需完善. 例如, 基于粗糙和模糊样本的结构风险最小化原则尚未建立. 再如, 第 6 章主要是对传统支持向量机的完善和推广, 尚未完全建立基于不确定结构风险最小化原则的支持向量机. 因此, 不确定统计学习理论仍有大量的工作需要开展.

我们于 2001 年开设了 “统计学习理论” 讨论班, 开始了对不确定统计学习理论的探讨, 感谢曾经参加或正在参加讨论班的杨兰珍、白云超、王鹏、唐文广、周彩丽、田静、李俊华、李嘉、李颜、王英新、张春琴、刘扬、田景峰、冯志芳、杜二玲、陈继强、郑莉芳、张新爱、马丽娟、孙璐、彭桂兵、闫舒静、张现坤、高林庆等博士生和硕士生, 他们的建议以及部分同学的硕士论文充实了本书的内容. 纪爱兵教授、李昆仑教授和邢红杰博士参加了本书的讨论, 提出了一些宝贵的意见, 并且第 6 章选用了他们关于支持向量机方面的部分研究成果, 在此一并致谢.

本书的部分研究内容得到了国家自然科学基金 (编号: 60573069, 60773062)、教育部科学技术研究重点项目(编号: 206012)、河北省自然科学基金 (编号: F2008000633)、

河北省教育厅科研计划重点项目 (编号: 2005001D) 和河北省高校学术著作出版基金的资助, 特此致谢.

由于作者学识和水平所限, 书中不足之处在所难免, 敬请同仁及读者批评指正.

作　者

2010 年 1 月 28 日

目　录

人们对于解决此类问题的努力一直在进行, 但是多数工作集中在对已有 (基于传统统计学原则的) 方法的改进和修正上, 或者学习某种启发式算法. Vapnik 等[2~6] 早在 20 世纪 60 年代就开始致力于研究小样本下的机器学习问题, 直到 90 年代中期, 小样本情况下的机器学习理论研究逐渐成熟起来, 形成了一个较完善的理论体系 —— 统计学习理论 (statistical learning theory, SLT)[7]. 该学习理论第一次强调了所谓小样本统计学习的问题. 研究表明, 对于很多函数的估计问题, 它可以得到比基于传统统计技术方法更好的解. 因此, 这一新的理论框架中的小样本统计学无论是在统计学习理论中还是在理论和应用统计学中, 都形成了一个前沿的研究方向. 它不仅为研究小样本下的统计模式识别和更广泛的机器学习问题建立了一个较好的理论框架, 同时以此为基础也发展了一种新的通用模式识别方法 —— 支持向量机 (support vector machine, SVM)[7,8]. 这种学习方法在解决小样本、非线性等问题中表现出许多特有的优势, 并能够推广到函数拟合等其他机器学习问题中 [7]. 在短短 40 余年的时间里, 国内外发表了大量此方面的论著[1~22]. 统计学习理论已经成为国际上机器学习领域继模式识别和神经网络之后新的研究热点[7]. 统计学习理论的主要内容包括以下 4 部分[7]:

(1) 经验风险最小化原则下学习过程一致性的条件 (重点是学习理论的关键定理);

(2) 在这些条件下关于统计学习方法推广性的界的结论 (重点是学习过程一致收敛速度的界、函数集的 VC 维及 VC 维基础上的推广性的界);

(3) 在这些界基础之上建立的小样本归纳推理原则 (主要是结构风险最小化原则);

(4) 实现新的结构风险最小化原则的实际算法 (重点是支持向量机).

(1)~(3) 是统计学习理论的基础理论部分, (4) 是基于统计学习理论的基础理论发展出的方法及应用部分.

1.2　不确定统计学习理论的提出和研究现状

目前, 统计学习理论被许多学者认为是处理小样本机器学习问题的最佳理论. 但伴随着对它的理论研究和实际应用的进一步拓广, 这一理论出现了自身难以解决的问题. 例如:

(1) 统计学习理论是建立在概率 (一类特殊的测度, 也可称为概率测度) 空间上的. 众所周知, 概率测度是一个满足可加性 (可列可加性) 的非负单值实数集函数. 由于可加性条件非常苛刻, 在实际应用中, 这个条件往往得不到满足, 换言之, 在实际应用中存在着大量的非可加集函数. 例如, 在测量问题中, 尽管经典测度的可加性可以很好地刻画许多类型的、理想的和无误差条件下的测量问题, 然而, 在测量

本统一称为广义不确定样本. 本书旨在构建统一的、处理广义不确定测度空间上基于广义不确定样本机器学习问题的统计学习理论 (简称为不确定统计学习理论). 不确定统计学习理论同样包括 4 部分主要内容: 广义不确定测度空间上基于广义不确定样本学习过程的一致性 (简称为不确定学习过程的一致性)、广义不确定测度空间上基于广义不确定样本的学习过程收敛速度的界 (简称为不确定学习过程收敛速度的界)、广义不确定测度空间上基于广义不确定样本结构风险最小化原则 (简称为不确定结构风险最小化原则)、广义不确定测度空间上基于广义不确定样本的支持向量机 (简称为不确定支持向量机). 不确定统计学习理论是传统统计学习理论的发展和拓广, 由于广义不确定测度是研究不确定统计学习理论的基础, 故在介绍不确定统计学习理论研究现状之前, 先扼要叙述几种除概率测度以外有代表性的广义不确定测度的发展概况如下:

(1) 定义于经典集合的若干子集组成的类, 并且取值为实数值的非可加测度研究. 1954 年, 法国数学家 Choquet[32] 提出了一种称为容度的理论. Choquet 容度是一个集函数, 它使得所设空间上的每一个子集均与一实数 (不要求非负) 对应, 它是连续且单调非减的. 受 Choquet 工作的启发, 1967 年, Dempster[33] 提出后经 Shafer 深化出来的两种类型的非可加测度, 分别称为信任测度与似然测度. 同时对这两种类型的非可加测度进行深入的研究, 便形成了 Dempster-Shafer 理论或显著性理论[34,35]. 1974 年, 日本学者 Sugeno[36] 在他的博士论文中首次提出了用比较弱的单调性和连续性来代替可加性的另一类集函数, 称之为模糊测度 (只是一个名称而已, 并无和模糊集对应的 "模糊" 含义). 国内外许多学者对此进行了研究, 得到了一些有意义的结果[37~43]. Liu[44] 于 2007 年提出基于正规性、单调性、自对偶性和可列次可加性的不确定测度, 并建立了基于不确定测度的不确定理论[44,45]. 2008 年, Wang 和 Klir[42] 建立了广义测度论. 2009 年, 哈明虎等[43] 提出了广义模糊集值测度, 并对其理论进行了初步研究.

(2) 定义在模糊集类上且取值为实值的测度研究. 1965 年, 美国控制论专家 Zadeh[26] 提出了模糊集的概念, 这标志着在众多领域有重要应用的新学科 —— 模糊数学的诞生, 从而导致了定义在模糊集上的测度的产生, 有人把这种测度也称为模糊测度. 近年来, 许多学者一直致力于这个方面的研究, 并取得了很多有意义的结果[37,46~49], 其中, 1978 年, Zadeh[47] 提出了给定模糊集合上的可能性分布函数 (possibility distribution function) 的概念, 并且定义了基于模糊集类的可能性测度 (possibility measure), 建立了可能性理论, 并将不确定性理解为可能性. 2002 年, Liu 和 Liu[50~52] 提出了可信性测度的概念, 并建立了公理化的可信性理论.

(3) 取值为集值或模糊值的测度研究. 集值测度是伴随着 20 世纪 40 年代集值映射的提出和发展而产生的[53~56]. 1972 年, Artstein[57] 率先在 $\mathbf{R}^n$ 空间中引入集值测度的概念, 得到了一些重要结果. 其后, 1978 年, Hiai[58] 对取值于 Banach 空

间的集值测度进行了讨论, 对有界变差的集值测度以空间的几何性质为工具建立了 Artstein 的相应结果. Papageorgiou[59] 进一步推广和补充了 Hiai 的工作, 给出了测度变换的表示定理. 张文修和李腾[60] 利用 Artstein 的选择定理讨论了由单值测度生成集值测度的条件. 薛小平[61] 利用不依赖于有限维特点的方法, 将上述有界闭凸值集值测度的表示定理推广到可分自反实 Banach 空间的情形, 并且给出了生成集值测度为有界变差的条件. Guo 和 Zhang[62] 于 2004 年定义了集值模糊测度.

自从模糊数的概念出现以后, 人们自然地考虑关于模糊数的测度问题. 1986 年, 张文修[63] 在 $\mathbf{R}^n$ 中定义了一种模糊数测度 (模糊集值测度), 给出了模糊数测度与集值测度的关系及模糊数测度的 Lebesgue 分解. 1998 年, Wu 等[64] 将模糊测度的值域推广到了模糊数.

下面再叙述不确定统计学习理论研究现状.

(1) 非概率测度空间上的统计学习理论研究. 哈明虎等[65,66] 将统计学习理论从概率测度空间扩展到 Sugeno 测度空间上, 提出了 Sugeno 测度空间上的经验风险最小化非平凡一致收敛的原则, 并给出了此空间上学习理论的关键定理和学习过程一致收敛速度的界, 也提出了 Sugeno 测度空间上带零均值噪声学习理论的关键定理; Bai 等 [67,68] 首先得到了可信性测度空间上学习理论的次关键定理, 继而又得到了可信性测度空间上的学习理论的关键定理和一致收敛速度的界; 哈明虎等[69,70] 在拟概率空间上得到了学习理论的关键定理和学习过程一致收敛速度的界; 哈明虎等[71,72] 在可能性空间上给出了学习理论的关键定理和学习过程一致收敛速度的界; 张现坤等[73,74] 也进一步讨论了不确定测度空间上的统计学习理论, 给出了不确定空间上学习理论的关键定理.

(2) 概率测度空间上基于带噪声、模糊、随机集等广义不确定样本的统计学习理论的基础理论研究. 哈明虎等[75,76] 给出了基于带零均值噪声样本的统计学习理论的关键定理, 讨论了基于带零均值噪声样本的学习过程一致收敛速度的界, 给出了受噪声影响的模糊样本学习理论的关键定理; Tian 等[77,78] 给出了基于模糊样本的学习理论的关键定理和学习过程一致收敛速度的界; Liu 等 [79] 给出了基于粗糙样本的学习理论的关键定理和学习过程一致收敛速度的界; 张植明等[80,81] 提出了基于复随机样本学习理论的关键定理、学习过程一致收敛速度的界、函数集的 VC 维、VC 维基础上推广性的界以及结构风险最小化原则; 孙璐等[82] 讨论了基于随机集样本学习理论基础. 此外, Ha 等[83] 讨论了基于模糊复随机样本的学习理论基础.

(3) 非概率测度空间上基于带噪声、模糊、随机集等广义不确定样本学习理论的基础理论研究. Ha 等[84] 提出了集值概率空间上基于随机集样本的学习理论, 给出了集值概率空间上基于随机集样本的关键定理、学习过程一致收敛速度的界, 并率先提出了随机集 VC 维的定义以及在此基础上的学习机器推广能力的界和结构

经验风险最小化原则. 王超[85] 给出了 Sugeno 测度空间上基于模糊样本的学习理论的关键定理和学习过程一致收敛速度的界. 张植明和田景峰[86] 给出了粗糙空间中基于双重粗糙样本的统计学习理论的理论基础.

(4) 不确定支持向量机的初步研究. Xing 等[87] 提出了一种新型的特征加权支持向量机. Takuya 和 Shigeo[88] 首次提出了模糊支持向量机的概念; Lin 和 Wang[89] 给出了基于隶属度的模糊支持向量机; 哈明虎等[90] 给出了一种新的模糊支持向量机, 改进了模糊支持向量机隶属度; 李昆仑等[91] 构建了模糊多类支持向量机并应用在入侵检测中; Ji 等[92] 提出了基于模糊训练数据的支持向量机. 尽管这些新的支持向量机也可以看成是不确定支持向量机的特例, 但是它们尚缺乏相应的不确定结构风险最小化原则的理论支撑, 而同样作为不确定支持向量机特例的传统支持向量机是基于结构风险最小化原则建立起来的. 杨志民和刘广利[93] 提出不确定性支持向量机, 但他们所指的不确定性支持向量机是根据模糊、粗糙、未确知理论和可能性测度等理论给出的, 虽然它们也可以纳入本书不确定支持向量机的范围, 但它们也不是根据相应的不确定统计学习理论中的不确定结构风险最小化原则构建的. 因此, 不确定统计学习理论仍有大量的工作去做.

参 考 文 献

[1] 白鹏, 张喜斌, 张斌等. 支持向量机理论及工程应用实例. 西安: 西安电子科技大学出版社, 2008

[2] Vapnik V N. The Nature of Statistical Learning Theory. New York: A Wiley-Interscience Publication, 1995

[3] 瓦普尼克. 统计学习理论的本质. 张学工译. 北京: 清华大学出版社, 2000

[4] Vapnik V N. Statistical Learning Theory. New York: A Wiley-Interscience Publication, 1998

[5] 瓦普尼克. 统计学习理论. 许建华, 张学工译. 北京: 清华大学出版社, 2004

[6] Vapnik V N. An overview of statistical learning theory. IEEE Transactions on Neural Networks, 1999, 10(5): 988~999

[7] 边肇祺, 张学工. 模式识别. 北京: 清华大学出版社, 1999

[8] Cortes C, Vapnik V N. Support vector networks. Machine Learning, 1995, 20: 273~297

[9] Cristianini N, Shawe-Taylor J. 支持向量机导论. 李国正, 王猛, 曾华军译. 北京: 电子工业出版社, 2004

[10] 邓乃扬, 田英杰. 数据挖掘中的新方法 —— 支持向量机. 北京: 科学出版社, 2004

[11] 邓乃扬, 田英杰. 支持向量机 —— 理论、算法与拓展. 北京: 科学出版社, 2009

[12] 张学工. 关于统计学习理论与支持向量机. 自动化学报, 2000, 26(1): 32~44

[13] Nello C, John S. An Introduction to Support Vector Machines. Cambridge: Cambridge University Press, 2000

[14] Osuna E, Freund R, Girosi F. An improved training algorithm for support vector machines. Neural Networks Processing VII-Proceedings of the 1997 IEEE Workshop, 1997: 276~285

[15] Platt J C. Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods-Support Vector Learning, 1999: 185~208

[16] Zhang X. Using class-center vectors to build support vector machines. IEEE Conference on Neural Networks for Signal Processing, 1999: 3~11

[17] 刘志刚, 李德仁, 秦前清等. 支持向量机在多类分类问题中的推广. 计算机工程与应用, 2004, 40 (7): 10~13

[18] 唐发明, 王仲东, 陈绵云. 支持向量机多类分类算法研究. 控制与决策, 2005, 20(7): 746~749

[19] Steinwart I. Consistency of support vector machines and other regularized kernel machines. IEEE Transactions on Information Theory, 2005, 51: 128~142

[20] Guermeur Y. VC theory of large margin multi-category classifiers. Journal of Machine Learning Research, 2007, 8: 2551~2594.

[21] Chapelle O, Sindhwani V, Keerthi S S. Optimization techniques for semi-supervised support vector machines. Journal of Machine Learning Research, 2008, 9: 203~233

[22] Tsang Y C, Zhang Y Q, Chawla N V, et al. SVMs modeling for highly imbalanced classification. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2009, 39(1): 281~288

[23] Aumann R J. Integrals of set-valued functions. Journal of Mathematical Analysis and Applications, 1965, 12(1):1~12

[24] 张文修. 集值测度与随机集. 西安: 西安交通大学出版社, 1989

[25] 张从军. 集值分析与经济运用. 北京: 科学出版社, 2004

[26] Zadeh L A. Fuzzy sets. Information and Control, 1965, 8: 338~353

[27] 张文修, 吴伟志, 梁吉业等. 粗糙集理论与方法. 北京: 科学出版社, 2001

[28] Puri M L, Ralescu D A. Fuzzy random variables. Journal of Mathematical Analysis and Applications, 1986, 114: 406~422

[29] Pawlak Z. Rough set. International Journal of Computer and Information Sciences, 1982, 11(5): 341~356.

[30] Klement E P, Puri L M, Ralescu D A. Limit theorems for fuzzy random variables. Proceedings of the Royal Society of London, Series A, 1986, 407: 171~182

[31] Kwakernaak H. Fuzzy random variables: definition and theorems. Information Sciences, 1978, 15: 1~29

[32] Choquet G. Theory of capacities. Annales de l' Institut Fourier, 1954, 5: 131~295

[33] Dempster A P. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics, 1967, 38(2): 325~339

度的对应关系的一个总括.

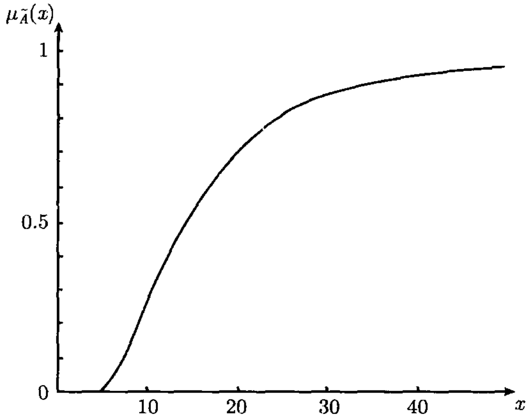

图 2.1 模糊集的隶属函数

例 2.1.2 设 $X=\{2,3,4,5,6,7\}$, $\tilde{A}$ 表示 "大约 5", 其隶属度如表 2.1 所示, 则 $\tilde{A}$ 可表示为

(1) $\tilde{A}=\dfrac{0}{2}+\dfrac{0.4}{3}+\dfrac{0.8}{4}+\dfrac{1}{5}+\dfrac{0.8}{6}+\dfrac{0.4}{7}$;

(2) $\tilde{A}=\{(2,0),(3,0.4),(4,0.8),(5,1),(6,0.8),(7,0.4)\}$;

(3) $\tilde{A}=(0,0.4,0.8,1,0.8,0.4)$.

表 2.1 模糊集的隶属度

x	2	3	4	5	6	7
$\mu_{\tilde{A}}(x)$	0	0.4	0.8	1	0.8	0.4

定义 2.1.2 设 $\tilde{A},\tilde{B}\in\tilde{\mathcal{F}}(X)$, 则

(1) 若 $\forall x\in X$ 恒有 $\mu_{\tilde{A}}(x)\geqslant\mu_{\tilde{B}}(x)$, 则称 $\tilde{A}$ 包含 $\tilde{B}$, 记为 $\tilde{A}\supseteq\tilde{B}$;

(2) 若 $\forall x\in X$ 恒有 $\mu_{\tilde{A}}(x)=\mu_{\tilde{B}}(x)$, 则称 $\tilde{A}$, $\tilde{B}$ 相等, 记为 $\tilde{A}=\tilde{B}$.

定义 2.1.3 设 $\tilde{A},\tilde{B}\in\tilde{\mathcal{F}}(X)$, 定义运算 $\tilde{A}\bigcup\tilde{B},\tilde{A}\bigcap\tilde{B},\tilde{A}^{c}$ 如下:

$$\mu_{\tilde{A}\cup\tilde{B}}(x)=\mu_{\tilde{A}}(x)\vee\mu_{\tilde{B}}(x),$$

$$\mu_{\tilde{A}\cap\tilde{B}}(x)=\mu_{\tilde{A}}(x)\wedge\mu_{\tilde{B}}(x),$$

$$\mu_{\tilde{A}^{c}}(x)=1-\mu_{\tilde{A}}(x).$$

$\tilde{A}\bigcup\tilde{B},\tilde{A}\bigcap\tilde{B},\tilde{A}^{c}$ 分别称为 $\tilde{A}$ 与 $\tilde{B}$ 的并集、交集和 $\tilde{A}$ 的补集.

例 2.1.3 设 $X=\{a,b,c,d,e\}$ 且

$$\tilde{A}=\frac{0.3}{a}+\frac{0.4}{b}+\frac{0.5}{c}+\frac{1}{d}+\frac{0.8}{e},\quad \tilde{B}=\frac{0.7}{a}+\frac{0.3}{b}+\frac{0.6}{c}+\frac{0.5}{d}+\frac{0.1}{e},$$

则

$$\begin{aligned}
\tilde{A}\bigcup\tilde{B}&=\frac{0.3\vee 0.7}{a}+\frac{0.4\vee 0.3}{b}+\frac{0.5\vee 0.6}{c}+\frac{1\vee 0.5}{d}+\frac{0.8\vee 0.1}{e}\\
&=\frac{0.7}{a}+\frac{0.4}{b}+\frac{0.6}{c}+\frac{1}{d}+\frac{0.8}{e},\\
\tilde{A}\bigcap\tilde{B}&=\frac{0.3\wedge 0.7}{a}+\frac{0.4\wedge 0.3}{b}+\frac{0.5\wedge 0.6}{c}+\frac{1\wedge 0.5}{d}+\frac{0.8\wedge 0.1}{e}\\
&=\frac{0.3}{a}+\frac{0.3}{b}+\frac{0.5}{c}+\frac{0.5}{d}+\frac{0.1}{e},\\
\tilde{A}^{\mathrm{c}}&=\frac{1-0.3}{a}+\frac{1-0.4}{b}+\frac{1-0.5}{c}+\frac{1-1}{d}+\frac{1-0.8}{e}\\
&=\frac{0.7}{a}+\frac{0.6}{b}+\frac{0.5}{c}+\frac{0}{d}+\frac{0.2}{e}.
\end{aligned}$$

模糊子集的并、交和补运算的性质如下：

(1) 幂等律：$\tilde{A}\bigcup\tilde{A}=\tilde{A},\tilde{A}\bigcap\tilde{A}=\tilde{A}$;

(2) 交换律：$\tilde{A}\bigcup\tilde{B}=\tilde{B}\bigcup\tilde{A},\tilde{A}\bigcap\tilde{B}=\tilde{B}\bigcap\tilde{A}$;

(3) 结合律：$(\tilde{A}\bigcup\tilde{B})\bigcup\tilde{C}=\tilde{A}\bigcup(\tilde{B}\bigcup\tilde{C}),(\tilde{A}\bigcap\tilde{B})\bigcap\tilde{C}=\tilde{A}\bigcap(\tilde{B}\bigcap\tilde{C})$;

(4) 吸收律：$\tilde{A}\bigcap(\tilde{A}\bigcup\tilde{B})=\tilde{A},\tilde{A}\bigcup(\tilde{A}\bigcap\tilde{B})=\tilde{A}$;

(5) 分配律：$(\tilde{A}\bigcup\tilde{B})\bigcap\tilde{C}=(\tilde{A}\bigcap\tilde{C})\bigcup(\tilde{B}\bigcap\tilde{C}),(\tilde{A}\bigcap\tilde{B})\bigcup\tilde{C}=(\tilde{A}\bigcup\tilde{C})\bigcap(\tilde{B}\bigcup\tilde{C})$;

(6) $\varnothing$ 与 X 满足：$X\bigcap\tilde{A}=\tilde{A},X\bigcup\tilde{A}=X,\varnothing\bigcap\tilde{A}=\varnothing,\varnothing\bigcup\tilde{A}=\tilde{A}$;

(7) 复原律：$(\tilde{A}^{\mathrm{c}})^{\mathrm{c}}=\tilde{A}$;

(8) 对偶律：$(\tilde{A}\bigcup\tilde{B})^{\mathrm{c}}=\tilde{A}^{\mathrm{c}}\bigcap\tilde{B}^{\mathrm{c}},(\tilde{A}\bigcap\tilde{B})^{\mathrm{c}}=\tilde{A}^{\mathrm{c}}\bigcup\tilde{B}^{\mathrm{c}}$.

为方便起见, 隶属函数 $\mu_{\tilde{A}}$ 简记为 $\tilde{A}$.

定义 2.1.4　设 $\tilde{A}\in\tilde{\mathcal{F}}(X)\ (\forall\lambda\in[0,1])$, 记

$$(\tilde{A})_\lambda=A_\lambda=\{x|\tilde{A}(x)\geqslant\lambda\},$$

称 A_λ 为 $\tilde{A}$ 的 λ- 截集, λ 为置信水平. 又记

$$(\tilde{A})_{\underset{\cdot}{\lambda}}=A_{\underset{\cdot}{\lambda}}=\{x|\tilde{A}(x)>\lambda\},$$

称 $A_{\underset{\cdot}{\lambda}}$ 为 $\tilde{A}$ 的 λ- 强截集. 显然,

$$\forall\lambda_1<\lambda_2\Rightarrow A_{\lambda_1}\supseteq A_{\lambda_2},A_{\underset{\cdot}{\lambda}_1}\supseteq A_{\underset{\cdot}{\lambda}_2}.$$

定义 2.1.5　设 $\tilde{A}\in\tilde{\mathcal{F}}(X)$, A_1 称为 $\tilde{A}$ 的核, 记为 $\mathrm{Ker}\tilde{A}$; $A_{\underset{\cdot}{0}}$ 称为 $\tilde{A}$ 的支集, 记为 $\mathrm{Supp}\tilde{A}$; $A_{\underset{\cdot}{0}}-A_1$ 称为 $\tilde{A}$ 的边界.

定理 2.1.1 (分解定理 1)　设 $\tilde{A}\in\tilde{\mathcal{F}}(X)$, 则 $\tilde{A}=\bigcup\limits_{\lambda\in[0,1]}\lambda A_\lambda$.

并且隶属函数为

$$f(\tilde{A}^{(1)}, \tilde{A}^{(2)}, \cdots, \tilde{A}^{(n)})(y) = \bigvee_{f(x_1,x_2,\cdots,x_n)=y} \left(\bigwedge_{k=1}^{n} \tilde{A}^{(k)}(x_k) \right),$$

$$f^{-1}(\tilde{B}^{(1)} \times \tilde{B}^{(2)} \times \cdots \times \tilde{B}^{(m)})(x) = \bigwedge_{j=1}^{m} \tilde{B}^{(j)}(y_j).$$

例 2.1.5 设 $X = \{1, 2, \cdots, 10\}$, 在 X 中定义模糊集如下:

$$\tilde{A} = \frac{1}{1} + \frac{0.8}{2} + \frac{0.6}{4}, \quad \tilde{B} = \frac{0.3}{2} + \frac{0.5}{3} + \frac{0.4}{5},$$

求 $\tilde{A} + \tilde{B}$.

解 根据扩展原理可得

$$\begin{aligned}
(\tilde{A} + \tilde{B})(3) &= \bigvee_{x_1+x_2=3} (\tilde{A}(1) \wedge \tilde{B}(2)) \\
&= 1 \wedge 0.3, \\
&= 0.3, \\
(\tilde{A} + \tilde{B})(4) &= \bigvee_{x_1+x_2=4} (\tilde{A}(x_1) \wedge \tilde{B}(x_2)) \\
&= (\tilde{A}(1) \wedge \tilde{B}(3)) \vee (\tilde{A}(2) \wedge \tilde{B}(2)) \\
&= (1 \wedge 0.5) \vee (0.8 \wedge 0.3) \\
&= 0.5.
\end{aligned}$$

同理,

$$(\tilde{A} + \tilde{B})(5) = 0.5, \quad (\tilde{A} + \tilde{B})(6) = 0.4, \quad (\tilde{A} + \tilde{B})(7) = 0.5, \quad (\tilde{A} + \tilde{B})(9) = 0.4,$$

$$(\tilde{A} + \tilde{B})(1) = (\tilde{A} + \tilde{B})(2) = (\tilde{A} + \tilde{B})(8) = 0,$$

即

$$\tilde{A} + \tilde{B} = \frac{0.3}{3} + \frac{0.5}{4} + \frac{0.5}{5} + \frac{0.4}{6} + \frac{0.5}{7} + \frac{0.4}{9}.$$

定义 2.1.9 设 $\tilde{A} \in \tilde{\mathcal{F}}(X)$.

(1) $\tilde{A}$ 称为凸模糊集当且仅当 $\forall \lambda \in [0, 1]$, A_λ 是凸集;

(2) $\tilde{A}$ 称为闭模糊集当且仅当 $\forall \lambda \in [0, 1]$, A_λ 是闭集, 即 $\forall x_n \in A_\lambda (n = 1, 2, \cdots)$, 并且 $\lim\limits_{n} x_n = a$, 则 $a \in A_\lambda$;

(3) $\tilde{A}$ 称为闭凸模糊集当且仅当 $\forall \lambda \in [0, 1]$, A_λ 是闭凸集;

(4) $\tilde{A}$ 称为正则模糊集当且仅当存在 $x_0 \in X$, 使得 $\tilde{A}(x_0) = 1$;

可知 x_1 能使 $\mathrm{e}^{-\left(\frac{x-a}{\sigma_1}\right)^2} \wedge \mathrm{e}^{-\left(\frac{z-x-b}{\sigma_2}\right)^2}$ 达到最大值, 从而

$$(\tilde{A}+\tilde{B})(z)=\mathrm{e}^{-\left(\frac{z-(a+b)}{\sigma_1+\sigma_2}\right)^2}.$$

注 2.1.2　$\tilde{A}(x)=\mathrm{e}^{-\left(\frac{x-a}{\sigma_1}\right)^2}$, $\tilde{B}(y)=\mathrm{e}^{-\left(\frac{y-b}{\sigma_2}\right)^2}$ 均为正态型模糊数.

注 2.1.3　设 $\tilde{a}\in\tilde{\mathbf{R}}^*$, 则根据模糊集的分解定理有

$$\tilde{a}=\bigcup_{\lambda\in[0,1]}\lambda a_\lambda=\bigcup_{\lambda\in[0,1]}\lambda[a_\lambda^-,a_\lambda^+].$$

定义 2.1.12　设 $\tilde{a},\tilde{b}\in\tilde{\mathbf{R}}^*$, 若 $\forall\lambda\in[0,1]$ 有

$$a_\lambda^-\leqslant b_\lambda^-,\quad a_\lambda^+\leqslant b_\lambda^+,$$

则称 $\tilde{a}\leqslant\tilde{b}$.

若 $\tilde{a}\leqslant\tilde{b}$ 且存在 $\lambda_0\in[0,1]$ 有

$$a_{\lambda_0}^-<b_{\lambda_0}^-\quad \text{或}\quad a_{\lambda_0}^+<b_{\lambda_0}^+,$$

则称 $\tilde{a}<\tilde{b}$.

若 $\tilde{a}\leqslant\tilde{b}$ 且 $\tilde{b}\leqslant\tilde{a}$, 则称 $\tilde{a}=\tilde{b}$.

定理 2.1.7　设 $\tilde{a},\tilde{b}\in\tilde{\mathbf{R}}^*$, 则 $\forall\lambda\in(0,1]$ 有

(1) $(\tilde{a}+\tilde{b})_\lambda=a_\lambda+b_\lambda=[a_\lambda^-+b_\lambda^-,\ a_\lambda^++b_\lambda^+]$;

(2) $(\tilde{a}-\tilde{b})_\lambda=a_\lambda-b_\lambda=[a_\lambda^--b_\lambda^-,\ a_\lambda^+-b_\lambda^+]$;

(3) $(\tilde{a}\cdot\tilde{b})_\lambda=a_\lambda\cdot b_\lambda=[c,\ d]$, 其中

$$\begin{aligned}c&=\min\left\{a_\lambda^-\cdot b_\lambda^-,a_\lambda^-\cdot b_\lambda^+,\ a_\lambda^+\cdot b_\lambda^-,a_\lambda^+\cdot b_\lambda^+\right\},\\ d&=\max\left\{a_\lambda^-\cdot b_\lambda^-,a_\lambda^-\cdot b_\lambda^+,\ a_\lambda^+\cdot b_\lambda^-,a_\lambda^+\cdot b_\lambda^+\right\}.\end{aligned}$$

2.1.2　粗糙集

粗糙集理论是波兰数学家 Pawlak 于 1982 年提出的, 它是一种新的处理模糊和不确定性知识的数学工具. 其主要思想是在保持分类能力不变的前提下, 通过知识约简导出问题的决策或分类规则, 详细内容参见文献 [4], [5].

若无特殊说明, 本书均设论域 U 为非空有限集合. U 上的一组划分称为关于 U 的一个知识库.

定义 2.1.13　假设 U 为非空有限集合 (也称论域), R 是定义在 U 上的等价关系, $[x]_R$ 表示包含元素 $x\in U$ 的 R 等价类, U/R 表示 R 的所有等价类的集合. 一个知识库就是一个关系系统 $K=(U,\mathbb{R})$, $\mathbb{R}$ 是 U 上的一族等价关系.