

基于云模型的规则提取算法研究

徐 飞¹,朱伟兴¹,李 众²

XU Fei¹,ZHU Wei-xing¹,LI Zhong²

1.江苏大学 电气信息工程学院,江苏 镇江 212013

2.常州信息职业技术学院,江苏 常州 213164

1.School of Electrical and Information Engineering,Jiangsu University,Zhenjiang,Jiangsu 212013,China

2.Changzhou Vocational College of Information Technology,Changzhou,Jiangsu 213164,China

XU Fei,ZHU Wei-xing,LI Zhong.Cloud model based rules extraction algorithm.Computer Engineering and Applications, 2008,44(2):63-66.

Abstract: A new scale cloud transform method is proposed firstly.Both the normal semi-cloud model and the shift operator are introduced to that method to make sure that data partition in the method can reflect some information about the actual data distribution,to reduce the number of rules generated by the new method,and to increase the rules' fitness.Because key points and best breakpoints are exploited,the rules base generated by the proposed method is complete.This paper first presents T-S cloud controller,and gives the simple design process of this new type of intelligent controller based on scale cloud transform.The simulation results show the new method is simple,powerful,and the T-S cloud controller consisted of the rules generated by scale cloud transform can reproduce the original data very well,and can reduce the control error to any degree theoretically.

Key words: qualitative reasoning;cloud model;rules extraction;cloud controller

摘 要:提出了尺度云变换的新方法。该方法采用“半正态云”作为基本云,并引入位移因子,使基于数据的规则提取算法中的数据划分能充分反映数据的实际分布,减少提取的规则数量,提高规则的推理精度;通过引入数据关键点及最佳分割点以保证规则库的完备性。还首次提出T-S型云控制器的概念,并将尺度云变换方法应用到T-S型云控制器的设计中。仿真结果表明,该方法简单有效,由该方法提取的规则构建的T-S型云控制器能较好地复现原数据分布,误差能控制在给定范围内。

关键词:定性推理;云模型;规则提取;云控制器

文章编号:1002-8331(2008)02-0063-04 **文献标识码:**A **中图分类号:**TP183;TP273

当前,智能控制应用日益广泛,取得的成果备受瞩目。然而,如何建立被控系统的知识模型是该领域进一步发展的瓶颈问题。对于简单的控制问题,可以直接将专家经验知识表述成“if-then”形式的定性推理规则;当问题较为复杂,控制系统的输入、输出变量维数较高时,专家的经验知识难以直接表达。在此情况下,利用被控系统的输入输出数据(样本数据)获取定性推理规则将是构建智能控制器的一个必要补充手段^[1]。此外,通过数据挖掘方法利用已有系统的样本数据来重构智能控制器,这也具有较高的学术意义和实用价值。

直接面向数据处理的规则提取算法^[1-5]都涉及输入和(或)输出空间的划分问题。如何合理有效地划分输入、输出空间,使建立的规则库在具有良好完备性和一致性的前提下,规则数量少,易于理解,且满足控制误差要求,一直是该领域的研究热点和重点。一般的划分方法主要存在以下两个问题:(1)硬性划分,即划分个数的确定带有主观性。不利于根据误差要求调整规则库。(2)划分不能反映数据的实际分布。例如对输入空间等距划分,导致低效划分,即不能同时保证规则的支持度和信任度,容易产生“catch-22”问题^[6]。

云变换^[7]是一种基于云模型^[8]的数据“软”划分方法,该方法可以将数据分布函数划分为若干个基于云模型表示的语言值,实现从定量数据到定性描述转换。

借鉴云变换在数据划分上的优越性,本文对云变换进行改进,提出尺度云变换方法,旨在探讨如何解决传统数据划分方法存在的两个问题。最后的仿真结果表明该方法简单、有效,且提取的规则满足控制误差要求。

1 云模型及云变换

1.1 云模型

定义 1^[8] 设 U 是一个普通集合, $U=\{x\}$ 称为论域, T 为 U 上的语言子集(语言值), $C_T(x)$ 是 U 到闭区间 $[0,1]$ 的映射,对于任意元素 $x \in U$,都存在一个有稳定倾向的随机数 $C_T(x)$,称式(1)为云模型。

$$C_T(x):U \rightarrow [0,1], \forall x \in U, x \rightarrow C_T(x) \quad (1)$$

云模型用数字特征 Ex 、 En 和 He 表示语言值,分别称之为期望、熵和超熵^[9]。图1为语言值“20岁左右”的云模型表示。如图1所示,在云模型中,定量值对定性值的隶属度是一组带有

稳定倾向的随机值,而不是一个确定值,这正反映了自然语言所固有的不确定性。云模型有效地实现了模糊性与随机性结合,解决了传统模糊理论存在的不彻底性^[7]。

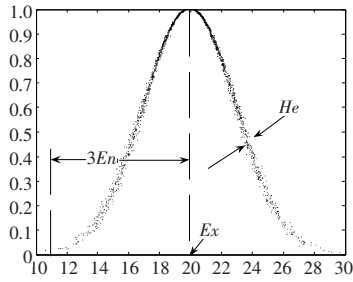


图1 定性语言值“20岁左右”的云模型表示

图2是用X条件云模型和Y条件云模型构造单规则定性推理结构^[9]。图中,CGx和CGy分别为X、Y条件云发生器,表示规则推理的前件和后件。由于用云模型表示的语言值集成了模糊性与随机性,基于云模型的定性推理也存在着固有的不确定性,即对应一个输入值x,CGx产生一组具有稳定倾向的随机输出 μ_i ,将输入空间的不确定性传递到输出空间;而CGy在同一个 μ_i 的多次刺激下也产生一组随机云滴 $drop(y_j, \mu_i)$ 。因为对输入空间中的一个值x,与其对应的输出空间值 y_j 具有不确定性,所以此推理结构实现了对不确定性的传播和继承。

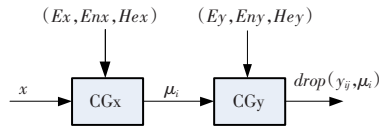


图2 一维云模型单规则推理结构示意图

文献[9]给出了一维云模型单规则推理结构及相应云控制器的结构设计及具体实现算法。

1.2 云变换

定义2^[7] 云变换是指对于任意一个不规则的数据分布,根据某种原则进行数学变换,使之成为若干个不同云的叠加,即满足

$$\left| g(x) - \sum_{i=1}^m c_i * \mu_i(x) \right| < e \quad (2)$$

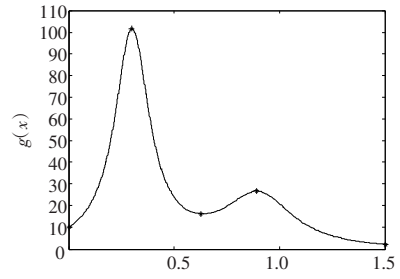
其中 $g(x)$ 为数据分布函数,令 $f_i(x) = c_i * \mu_i(x)$ 为基于云模型表示的语言值的数据分布函数, $\mu_i(x)$ 为云模型的期望曲线^[10],用数字特征 (Ex_i, En_i) 确定; c_i 为尺度变换因子, m 为叠加的云的个数, e 为误差上限。

将用于上述变换的云模型称为基本云,简称基云。从实际数据分布中提取若干个基于云模型的语言值的过程,实质是用云模型来拟合数据分布函数的过程,即对式(2)各项参数的求解过程。

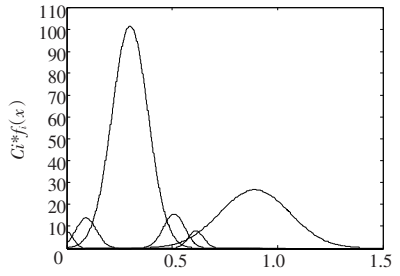
设数据分布函数为 $g(x)$,如图3(a)所示。选用正态云模型作为基云,图3(b)、(c)分别为各语言值的数据分布函数及云模型表示。从数据挖掘和知识发现的角度看,云变换就是从某一论域的数据分布中恢复其定性描述的过程,是从定量描述到定性描述的转换过程,也是一个概念归纳学习的最优化过程^[7,10,11]。

2 尺度云变换及T-S型云控制器

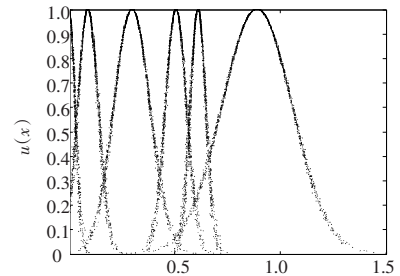
为了将云变换方法应用到定性规则的提取中,本文提出尺



(a)数据分布函数 $g(x)$



(b) $e=5$ 时语言值的数据分布函数 $f_i(x)$



(c)语言值的云模型表示

图3 云变换示意图

度云变换方法。

2.1 尺度云变换

定义3 尺度云变换是指对任意一个数据分布 $g(x)$,根据其关键点进行有限次划分,并对各个分区内的数据进行数学变换,包括位移变换和尺度变换,使得每一个区间数据仅用一条云模型规则描述,即

$$g(x) \approx \sum_{i=1}^m (c_i * \mu_i(x) + s_i)$$

且上式满足:

$$\left| g(x) - \sum_{i=1}^m (c_i * \mu_i(x) + s_i) \right| < e$$

令 $f_i(x) = c_i * \mu_i(x) + s_i$ 为第 i 个云模型规则的数据分布函数, m 为分区个数, m 与误差上限 e 有关, $\mu_i(x)$ 为第 i 个云模型的期望曲线,用云模型数字特征 (Ex_i, En_i) 确定, c_i 为尺度变换因子, s_i 为位移因子。

尺度云变换方法采用“半正态云”^[10]作为基云。仍以数据分布函数 $g(x)$ 为例,则尺度云变换的算法描述如下:

步骤1 求取控制曲线(面)关键点,包括控制曲线(面)的起始点和终止点,用 $K_i(i=1,2,\dots,m+1)$ 表示。

步骤2 相邻关键点 $K_N, K_{N+1}(N$ 为整数, $1 \leq N \leq m)$ 间建立一条基于云模型的定性规则 R_N ,具体过程为:

(1)确定位移因子和尺度变换因子:

$$s_N = \min(|\lg(K_N)|, |\lg(K_{N+1})|)$$

$$c_N = |g(K_N) - g(K_{N+1})|$$

图4中,控制关键点“*”将 $g(x)$ 的划分为四个区间,对数据段 a 进行位移变换,即 $g(x) - s_1$,得到 a' 。

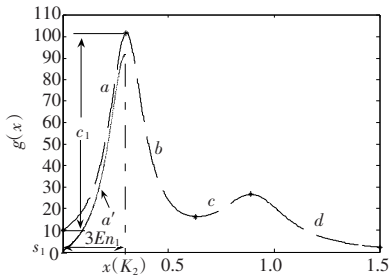


图4 位移因子 s_1 和尺度变换因子 c_1

(2)确定 $\mu_N(x)$ 的参数:

$$Ex_N = \begin{cases} x(K_N), g(K_N) > g(K_{N+1}) \\ x(K_{N+1}), \text{others} \end{cases}$$

$$En_N = \frac{x(K_{N+1}) - x(K_N)}{3}$$

则规则 R_N 的云模型的期望曲线为:

$$\mu_N(x) = e^{-\frac{(x-Ex)^2}{2En^2}}, x \in [x(K_N), x(K_{N+1})]$$

关于超熵 He_N ,本文采用如下确定方法:

$$He_N = \begin{cases} \frac{1}{10k}, k > 100 \\ 0.005, \text{others} \end{cases}$$

其中 $k = \max(g(K_N), g(K_{N+1}))$ 。超熵的确定应保证输入值对规则的隶属度的随机性在一个合理的范围内。 He_N 可以根据实际数据分布的变化及误差要求作相应地调整,灵活性较大。用语言值 A_N 表示区间 $[x(K_N), x(K_{N+1})]$,这样就得到了一条定性规则:

if x is A_N , then $y = c_N * \mu_N(x) + s_N$

步骤3 测试规则 R_N 的误差 e_N ,即

$e_N = \max(|g(x) - c_N * \mu_N(x) - s_N|)$,且 $x \in [x(K_N), x(K_{N+1})]$ 。若 $e_N < e$,跳转至步骤2,提取描述下一个数据区间的规则;否则对数据区间 $[x(K_N), x(K_{N+1})]$ 引入最佳分割点 T ,分别对区间 $[x(K_N), x(T)]$ 和 $[x(T), x(K_{N+1})]$ 进行步骤2操作,产生两条新的规则 R_{N1} 和 R_{N2} ,并测试 R_{N1}, R_{N2} 的误差,若不满足误差要求,则循环执行步骤2和步骤3,直至产生的规则满足误差要求。最佳分割点 T 使得 R_{N1} 和 R_{N2} 的误差最小,即 $e_{N1}^2 + e_{N2}^2$ 最小。

本文采用二分法寻找最佳分割点 T ,图5中,从数据段 a 提取的半云模型规则的数据分布函数为:

$$f_1(x) = c_1 * \mu_1(x) + s_1$$

$$e_1 = \max(|g(x) - f_1(x)|)$$

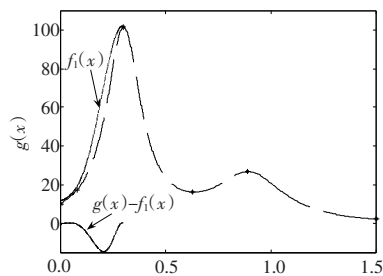


图5 规则 R_1 的误差分析

若 $e_1 > e$,则对数据段 a 进行二次划分,并对两个新区间分别进行步骤2、步骤3操作;否则,直接进行步骤2操作,提取描述下一个区间数据的规则。

循环执行上述3个步骤,直至对所有数据段的尺度云变换完成。

2.2 规则提取及结果分析

为了验证尺度云变换方法应用于定性规则提取的可行性,进行如下仿真实验:

(1)实验参数

如图1所示,输入输出数据对 (x_i, y_i) , $(i=1, 2, \dots, 1000)$,
 $x_i \in [0, 1.5], y_i = \frac{1}{(x_i - 0.3)^2 + 0.01} + \frac{1}{(x_i - 0.9)^2 + 0.04} - 1$,误差上限 $e=5$ 。

(2)实验结果及分析

对上述I/O数据提取定性控制规则,得到结果为:数据划分为7个区间,表1为各规则的参数。

表1 规则参数

R_i	μ_i			c_i	s_i	e_i	L/R
	Ex	En	He				
1	0.077	0.026	0.005	6.997	10.177	0.859	L
2	0.300	0.074	0.001	84.328	17.173	3.336	L
3	0.300	0.072	0.001	79.543	21.956	2.944	R
4	0.516	0.037	0.005	5.679	16.277	1.224	R
5	0.888	0.087	0.005	10.446	16.277	0.520	L
6	0.888	0.166	0.005	23.265	3.462	1.195	R
7	1.385	0.038	0.005	1.273	2.190	0.185	R

注:误差上限 $e=5$,L/R 分别表示左、右半正态云。

图6为规则的云模型数据分布与I/O数据拟合比较及拟合误差。“*”表示对I/O数据的初次划分,“+”表示二次划分的最佳分割点。图7为各规则的云模型表示。

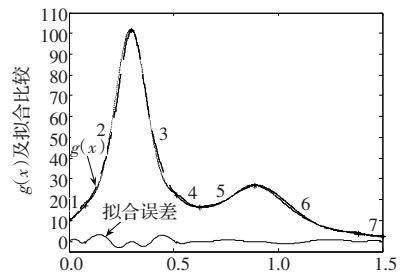


图6 拟合比较及拟合误差曲线

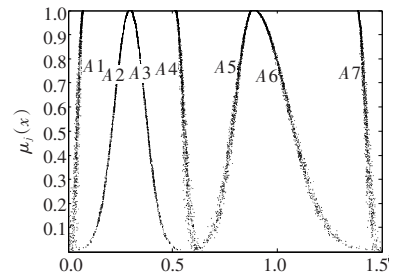


图7 规则的云模型表示

分析图3、图6和图7,得出尺度云变换方法特点如下:(1)划分体现数据的分布且反映数据变化趋势,即规则的可理解性较好。图7中,云模型反映原数据的信息,包括控制关键点及数据的变化趋势,而图3(c)却不能反映出这些数据信息。(2)由尺度云变换算法可知,由于引入关键点及最佳分割点,使得到

的规则库具有较好的完备性。若数据分布函数 $g(x)$ 的最小值小于误差上限 e , 则云变换方法不能保证规则库的完备性。图 3 中, $x=1.5$ 时, $g(x)$ 的最小值 $g(1.5)=2.19$, 若误差上限 $e=5$, 则 $x=1.5$ 处无可用规则, 即规则库不具有完备性。由于尺度云变换方法是对分区内的数据进行总体描述, 保证了规则库的完备性。(3) 规则库中规则数目较少, 且与原数据的拟合误差较小。事实上, 在数据分布函数关键点领域内, 数据分布一般是不对称的, 故采用半正态云作为基云, 以减少拟合误差; 引入位移因子, 通过该线性算子对数据进行预处理, 使得单条规则的拟合误差较小。此外, 尺度云变换采用二次最佳划分方法减少误差, 这样仅需增加少量的二次划分就可使误差迅速减小, 使得规则的数量较少。(4) 规则库可维护性较好。首先, 可以根据误差要求调整规则库, 例如, 若提高误差要求, 使 $e=3$, 由表 1 可知, 只需对图 6 中的数据段 2 进行二次划分, 重新提取规则即可; 其次, 可以根据数据的变化调整规则库, 例如, 若数据段 5 的 I/O 映射关系改变, 仅需调整 R_5 的参数。这样, 尺度变换方法就保证了所得到的规则库具有良好的可维护性。

2.3 T-S 型云控制器

为了将尺度云变换得到的规则应用于定性控制, 借鉴 T-S 模糊控制器的原理, 本文首次提出 T-S 型云控制器, 其控制规则为:

RN: if x is A_N , then $y=f_N(u_N(x))$

规则的推理结构为图 8 所示:

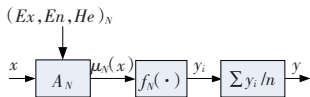


图 8 T-S 云控制器单规则推理结构

图 8 中, A_N 为定性规则的前件, 是用云的三个数字特征表示的 x 条件云发生器, 这样, 输入值 x 多次刺激 A_N 时, 其隶属度 $\mu_N(x)$ 是一组有稳定倾向的随机数, 经过线性算子 $f_N(\cdot)$ 得到一组输出值 y_i , 取其平均值作为最终的输出。由此可构建 N 规则 T-S 型云控制器, 如图 9 所示。

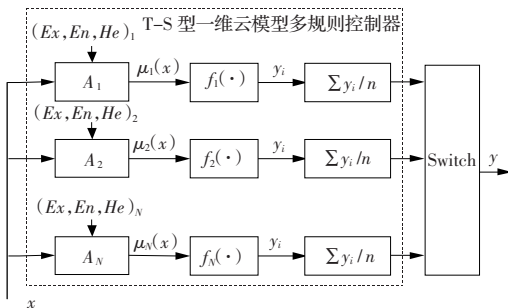


图 9 T-S 型云控制器

为验证上述规则提取算法, 设计以下实验: 在 MATLAB 平台上实现 T-S 型云控制器, 并封装表 1 中的规则, 控制器的输入值满足 $x_i \in [0, 1.5] (i=1, 2, \dots, 1000)$, 比较控制器输出值 y'_i 和原 y_i 。图 10 为比较结果及误差分析。求得误差的最大值为 4.7173, 小于设定的误差上限, 满足误差要求。图 10 中, 原控制曲线与控制器输出曲线几乎重合在一起。

3 结论

本文对云变换进行了改进, 提出尺度云变换方法。新方法

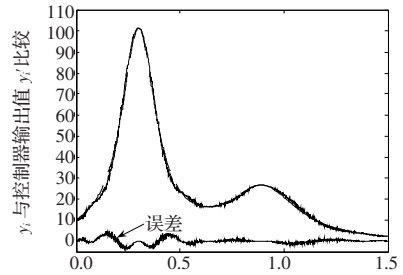


图 10 控制器输出与原数据比较分析

采用半正态云作为基云, 并引入位移因子, 以减少规则的数量并提高推理精度; 引入关键点及最佳分割点以保证所得到的规则库的完备性。运用尺度云变换方法对数据进行划分, 其划分结果体现实际数据的分布, 得到的规则库具有良好的完备性及可维护性, 规则数目少且易于理解。

本文还首次提出 T-S 云控制器的概念, 并在此基础上验证了尺度云变换方法提取规则的有效性。最后的仿真实验表明, 用尺度云变换方法得到的规则构建的 T-S 型云控制器较好地复现了原数据分布, 这说明该方法能较好地输入输出数据中提取定性知识, 实现定量描述到定性描述的转换。

文中仅讨论了单值函数数据的情况, 今后的工作将探讨多值函数数据的规则提取算法, 此外, 对尺度云变换方法进行改进使之适用于高维输入输出数据的处理将是今后研究工作的另一个重点。(收稿日期: 2007 年 7 月)

参考文献:

- [1] Wang Li-xin, Mendel J M. Generating fuzzy rules by learning from examples[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1992, 22(6): 1414-1427.
- [2] Abe S, Lan M. Fuzzy rules extraction directly from numerical data for function approximation[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1995, 25: 119-129.
- [3] Abe S, Lan M. A method for fuzzy rules extraction directly from numerical data and its application to pattern classification[J]. IEEE Transactions on Fuzzy Systems, 1995, 3: 18-28.
- [4] Mikhailov L, Nabouh A, Lekova A. Method for fuzzy rules extraction from numerical data[C]//Proceedings of the 12th IEEE International Symposium on Intelligent Control, Istanbul, Turkey, 16-18 July 1997: 61-65.
- [5] 吴涛, 许晓鸣, 方兴其. 一类提取模糊规则的新方法及其在干燥建模中的应用[J]. 控制与决策, 2000, 15(2): 177-180.
- [6] Srikant R, Agrawal R. Mining quantitative association rules in large relational table[C]//Jagadish H V, Mumick I S. Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data Montreal, Canada: ACM Press, 1996: 1-12.
- [7] 杜鹤, 李德毅. 基于云的概念划分及其在关联采掘上的应用[J]. 软件学报, 2001, 12(2): 196-203.
- [8] 李德毅, 史雪梅, 孟海军. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(5): 15-20.
- [9] 高建, 李众. 一维云模型映射器设计及其应用研究[J]. 系统仿真学报, 2006, 18(7): 1861-1865.
- [10] 李德毅, 杜鹤. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005.
- [11] 王树良. 基于数据场与云模型的空间数据挖掘和知识发现[D]. 武汉: 武汉大学, 2002.