

基于大型模糊认知图的复杂系统建模与推理研究

彭 珍^{1,2} 田立勤^{1,2} 吴 静^{1,2} 高晓燕² 杨炳儒³

(青海师范大学计算机学院 青海 810008)¹ (华北科技学院计算机学院 北京 101601)²

(北京科技大学计算机与通信工程学院 北京 100083)³

摘 要 模糊认知图作为一种智能计算工具,具有直观的知识表达、快速的数值推理能力等优点,适用于系统建模与推理。为了实现复杂系统关联认知与聚类的集成挖掘,进而达到对它的有效分析与决策,在面向复杂系统的大型模糊认知图研究基础之上,提出了用于三江源生态决策的大型模糊认知图关联认知、状态建模与推理的思路与方法,它不仅可以提升大型模糊认知图在复杂系统建模与推理方面的理论研究,也将拓展其应用领域。

关键词 复杂系统,模糊认知图,集成挖掘,三江源生态系统

中图分类号 TP182 **文献标识码** A

Research on Complex System Modeling and Reasoning Based on Large Fuzzy Cognitive Map

PENG Zhen^{1,2} TIAN Li-qin^{1,2} WU Jing^{1,2} GAO Xiao-yan² YANG Bing-ru³

(School of Computer, Qinghai Normal University, Qinghai 810008, China)¹

(School of Computer, North China Institute of Science and Technology, Beijing 101601, China)²

(School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)³

Abstract As an intelligent computational tool, Fuzzy Cognitive Map has the advantage of intuitive knowledge representation and fast numerical reasoning ability, etc., and is applied for system modeling and reasoning. In order to achieve the integrated mining of associated cognitive and clustering, and to achieve effective analysis and decision-making for complex system, large Fuzzy Cognitive Map of complex system was researched, and the ideas and methods of modeling and reasoning of large Fuzzy Cognitive Map association cognitive and state were proposed for Three Rivers ecological decision, which will not only enhance the large Fuzzy Cognitive Map research on complex systems modeling and reasoning, but also expand its application area.

Keywords Complex system, Fuzzy cognitive map, Integrated mining, Three rivers ecosystem

1 引言

1986年Kosko在Axelord认知图的基础上,把概念间具有的三值逻辑关系扩展到 $[-1, 1]$ 区间的模糊关系上,提出了模糊认知图(Fuzzy Cognitive Map, FCM)^[1],如图1所示。FCM是一个将模糊反馈动力系统因果事件、参与值、目标与趋势通过各概念结点间的有向弧连接起来的图结构,每个结点都有一个状态空间,FCM状态空间经阈值/演变函数自动传播,通过整个网络各概念结点间的相互作用模拟系统的动态行为。由于FCM直观的知识表达、快速的数值推理能力及其与神经网络、图论、模糊逻辑等领域的密切关系,使其已广泛应用于政治、经济、工程、医学等各个领域^[2-5]。针对传统FCM在解决复杂系统建模问题中的缺陷,大型FCM对传统FCM进行了发展与创新。大型FCM体现了多个FCM的融合,融合可能是多维的,或多层次的,或联合的等,总之应是对多个个体的整体表现,是对个体总和的涌现;而且,大型

FCM中结点之间的相互作用仍为一种非线性关系。

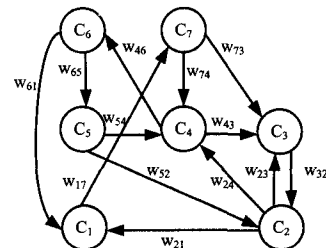


图1 FCM结构

总之,对大型FCM的研究已成为FCM的发展趋势与紧迫任务之一。大型FCM是一种有效的复杂系统建模方法,对它的研究已取得了一定进展。基于此,本文以三江源生态监测数据为驱动,拓展与提升大型FCM在复杂系统建模与推理方面的研究,实现对三江源生态系统的关联认知与聚类的集成挖掘,进而达到对它的有效分析与决策。

到稿日期:2012-09-25 返修日期:2012-12-11 本文受973计划项目(2011CB311809),省自然科学基金(2012-Z-932Q, 2012-Z-935Q, 2012-N-525),中央高校基本科研业务费专项(JSJ2013B07, 3142013098),国家自然科学基金(61175048)资助。

彭 珍(1981—),女,博士,副教授,主要研究方向为知识发现, E-mail: zhen_peng1981@163.com; 田立勤(1970—),男,博士后,教授,主要研究方向为智能网络、三江源生态监测; 吴 静(1980—),女,博士,讲师,主要研究方向为智能网络、三江源生态监测; 高晓燕(1970—),女,博士,副教授,主要研究方向为智能网络; 杨炳儒(1946—),男,博士生导师,主要研究方向为知识发现、智能系统。

2 面向复杂系统的大型 FCM 研究

目前,大型 FCM 实现复杂系统建模的总体思路是分而治之,即用 FCM 模拟复杂系统中的各子系统,再对它们进行融合。主要包括以下 3 种情况:

(1)在各个 FCM 的上层建立一个高层 FCM,从整体上构成一个多层次结构的 FCM。

文献[6]提出了直接从关系数据库中构造二层树型 FCM 的方法,即为单个表/关系构建一个对应的 FCM,高层构建一个模糊认知超图,它表示了不同表之间的关系。文献[7]提出对系统中的概念结点进行分组,在组结点上构造商 FCM,商 FCM 表示了系统的整体信息,商结点状态其实就是对局部 FCM 的状态融合,文章中只是提供了最大值、最小值、算术平均值、几何平均值或加权值等几种求商结点状态的方法,但这些方法都缺少具体的分析与论证。文献[8]提出以数据为驱动使用互信息与梯度下降法来实现高层结点状态的算法,并将其应用于文本分类中。

(2)各个 FCM 的结点相同,但是结点之间的关系不同,通过融合结点之间的关联结构与关联度形成一个大型 FCM。

文献[9]提出了一种基于 D-S 证据理论的融合方法,其将概念结点间因果关系的影响程度的可能取值构成识别框架,在专家识别框架上将某个权值评估的模糊值的隶属度作为信度分配函数,合成后的信度分配函数作为最终权值合成的依据,但这种方法仅建立在专家知识的基础上。文献[10]采用了一种基于遗传算法的计算方式,从宏观上为每个 FCM 设置一个融合信度,根据可信度实现多个 FCM 的融合,可信度是由它的样内误差来衡量,样内误差越小则可信度越大。

(3)各个 FCM 的结点不全相同,通过融合相同结点之间的关联结构与关联度形成一个大型 FCM。

文献[11]提出在各 FCM 合成时,相同概念结点间的关联权重需要进行综合/融合,选用基于信任知识库的信任度且概率最大的权重,当概率相近无法选取时,采用心智结点达到最佳的方法选取权重,但这种方法的前提是必须有可靠的信任知识库。

3 三江源生态系统的大型 FCM 建模研究

3.1 三江源生态系统的 FCM 结构

以三江源生态环境传感器监测数据为驱动,监测情况如表 1 所列,表中每一地区的监测内容即是监测类型,如气象监测、土壤监测与水文水质监测等;每种监测类型下又涉及到多个监测指标。为区域中每种监测类型创建一个 FCM,用于模拟其中监测指标之间的关系,FCM 中的概念结点表示监测指标对象,概念结点的状态可以是监测指标的多维模糊值,概念之间的关联表示监测指标之间的因果关联关系;在此基础上,建立高层 FCM,它反映了各种监测类型对象之间的关系,其中的结点代表由其监测指标对象组成的 FCM;根据具体情况层次可继续拓展;相邻层次之间建立影响与反馈关联关系,从而构造多维、多层次 FCM 的静态空间结构,如图 2 所示。

FCM 通过在不同时间上的结点状态及其演化来体现系统的行为,这种非线性演变体现在结点状态在不同时刻/步骤的转变中,其演变/阈值函数如式(1)所示。结点状态 $A_j(t+1)$

1)既与自身状态 $A_j(t)$ 有关,又受同层相关结点状态 $A_k(t)$ 的影响,还与其所在的上层相关结点 $A^i(t)$ 及下层结点 $A_{ji}(t)$ 有关。

$$A_j(t+1) = f(\alpha A_j(t) + \beta \sum_{j \neq k} A_k(t) w_{kj} + \gamma g_j(A^i(t)) + \lambda \varphi(A_{ji}(t))) \quad (1)$$

在此基础上,可由专家知识,或从数据集中计算,或两者结合的方法初始化 FCM 的关联矩阵,获得 FCM 初始结构。

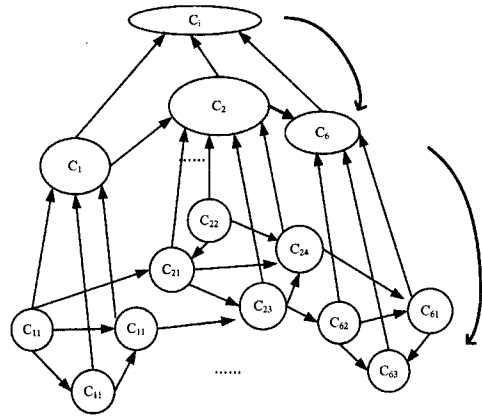


图 2 大型多层次 FCM 结构

表 1 部分传感器监测列表

监测类型	监测指标
气象监测	空气温度 at
	空气湿度 ah
	大气压 ap
	总辐射 ar
	大气 CO ₂ 含量 ac
土壤监测	土壤温度 st
	土壤含水量 sw
	土壤热通量 sh
	土壤 CO ₂ 含量 sc
水文水质监测	水体温度 ht
	电导率 hd
	水体 PH 值 hp
	溶解氧含量 ho

3.2 三江源生态系统的大型 FCM 结点状态

FCM 中结点状态的获知是其关联挖掘的基础,因为在关联优化中需要获知每个结点的状态值。而复杂系统的大型 FCM 中高层结点是对其底层对应结点的聚集,因此,对于大型 FCM 的挖掘,即为关联认知与聚类的集成挖掘[12]。

在三江源生态监测数据中,特征层结点状态来自多传感器监测的数据特征,是多维的,每维特征数据看作一个输入结点,特征层结点看作输出结点,从而形成一个多输入、单输出的局部简单 FCM。

高层中不同层次代表了不同的决策级别,高层结点的融合状态是由该结点对应的下层 FCM 与该结点间相互作用而来,不同决策级别的状态融合是对其底层结点状态的聚类,由此可以实现三江源不同级别的生态健康状况评价。

聚类方法可采用非线性 Hebbian 学习方法、或以非线性 Hebbian 学习为核心的方法[13,14],在该算法中 FCM 演变/阈值函数的选择很重要,选用累计正态分布函数或者近似的累积逻辑分布函数,如双曲正切函数 tanh,可使输出结果呈现出 U 型分布[15],从而涌现得到高层结点状态。根据 NHL 学习的特征,设定 NHL 学习的结束条件(目标函数等)。根据矩

阵范数 $\|w\|=1$ 的限制,调整结点间关联权值,直到满足结束条件为止。

状态的稳定性可采用 Lyapunov 理论分析证明。从动力学系统观点来看,稳定状态可以理解为系统某种形式的能量函数在系统运动过程中其能量值不断减少,最后趋于目标。这里的能量函数表征了 FCM 状态的变化趋势。由此,根据 NHL 学习的目标函数,定义系统的 Lyapunov 函数 $E(w)$,通过证明 $E(w)<0$,且 $\frac{dE}{dw}\geq 0$,得到 $E(w)$ 是系统的 Lyapunov 函数,且系统在目标函数处趋于稳定。

3.3 三江源生态系统的大型 FCM 关联优化

在三江源生态系统中,大型 FCM 的关联认知度的优化可以获得监测对象间的直接关联认知度。

监测的特征数据体现了低层结点状态,融合状态反映了高层结点的状态,特征数据的变化反映了结点状态的变化。可采用进化算法来实现关联优化,以遗传算法^[16]为例,将染色体定义为 FCM 中的所有关联集,经选择、交叉、变异操作,将下一时刻/步骤来自数据源中的实际状态值与 FCM 计算出的下一时刻/步骤的状态值进行对比,以此训练关联矩阵,形成新一代群体,直到达到最大适应度或最大代数。

为提高建模算法的效率,可对大型 FCM 进行分解和组合,使用主从模式的并行机制来提高挖掘算法的效率。按照不同监测类型将其分解为若干个 FCM,每一个 FCM 由一个从进程负责计算状态和适应度评估,然后由主进程进行组合完成高层 FCM 的关联挖掘(如图 3 所示)。最后进行算法时间复杂度的分析、对比与评价。

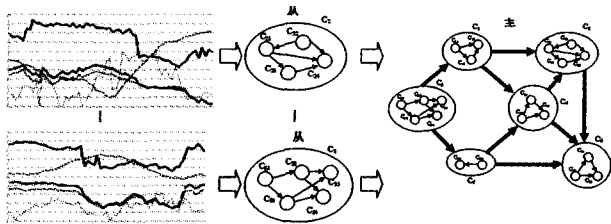


图 3 大型 FCM 并行计算示意图

为了提高建模算法的准确性,可加入平衡性、关联约束机制。平衡性是 FCM 中冲突评估的一种方法,平衡性由平衡度测定,定义路径的平衡度计算方法,由每条路径的平衡度构造 FCM 的平衡度^[17],将平衡度的计算融入适应度函数中,显然平衡度越高说明 FCM 越平衡,说明这个关联权集的适应度越高。当结点状态发生变化时,可通过对与结点状态转换有着密切关系的演化函数求导进行分析,根据状态增大/减少时因果关联程度的取决因素及变化,产生关联约束机制,并融入适应度评估中。平衡性与关联约束机制同时也克服了建模算法的盲目性,有利于提高算法的运行效率。

4 三江源生态系统的大型 FCM 推理研究

4.1 三江源生态系统的大型 FCM 关联认知推理

在三江源生态系统中,大型 FCM 中结点之间关联程度的推理是为获得监测对象间的间接关联认知度。

首先,从结点之间的复杂关系入手,定义结点之间的关联路径,根据结点间的层次与隶属关系,以及直接/间接关系,分析结点之间的间接关系。然后,定性分析,即根据 FCM 中因果关系、隶属关系权值变化的性质,通过结点间路径经过的

正/负有向弧/关联的个数,可分析结点之间的影响程度是正还是负,增加还是减少。最后,定量分析,即拟通过分析结点间间接关联的影响要素建立其间接影响关系,以图 4 为例,结点 C_i 对 C_k 在 t 时刻的间接影响与此路径上的影响程度 w_{ij} 、 w_{jk} 有关,还与间接路径中经过的结点 C_j 在 $t-1$ 时刻的关联有关,两结点间的影响程度是两结点间多条关联路径中影响程度的综合。

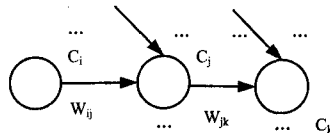


图 4 C_i 对 C_k 的间接关联

4.2 三江源生态系统的大型 FCM 状态变化推理

在三江源生态系统中,大型 FCM 状态变化的推理包括结点状态变化的主要因素与响应程度推理、图动态行为演变的挖掘,由此可以获得对象状态变化的主要因素与响应程度,以及生态的演变趋势。

结点状态变化的影响体现在该结点所在关联路径上的其他结点状态上,因此,从 FCM 中结点状态(特征)层面推理结点状态变化的主要因素与响应程度的方法是在以该结点为结果的关联路径中按照由近到远,由多到少的原则,在关联推理的基础上,根据结点之间直接与间接的关联性/关联程度,定性/定量分析该结点状态变化的主要影响因素。同样,在以该结点为原因的关联路径中分析结点状态变化引起结点状态变化的响应程度。

从 FCM 状态(决策)层面,推理 FCM 动态行为演变趋势。动态行为演变趋势体现了图中所有结点状态随时间的变化趋势,从关联路径入手以马尔可夫链为主要方法,重点分析以下两大方面:第一是 FCM 的状态,包括初始状态,比如初始状态全为零则演化趋势必定是固定点;包括结点状态转换次数,它与终止状态也存在一定的关系,比如在一定阈值函数作用下 FCM 偶数次结点状态转换才可能生成有限环,还有有限环的长度不大于状态转换的结点个数等;包括某时刻结点状态变化,比如若某时刻/步骤只有一个结点状态发生变换,研究下一步/时刻结点状态的变化等。第二是 FCM 的阈值/演变函数,包括函数类型,比如二值函数、三值函数、S 型函数;包括函数阈值,比如阈值/演变函数的取值范围对状态有很大影响,因为 FCM 的状态是通过阈值/演变函数进行转换的。

结束语 在充分研究大型 FCM 的基础上,针对三江源生态监测现状,提出使用大型层次 FCM 对三江源生态复杂系统进行建模与推理的思路与方法。对该大型 FCM 的集成(关联与聚类)挖掘可以实现三江源不同级别的生态健康状况评价、三江源生态监测对象间的直接与间接关联认知度挖掘、三江源生态对象状态变化的主要因素与响应程度以及生态的演变趋势分析。

参考文献

- [1] Aguilar J. A Survey about Fuzzy Cognitive Maps Papers (Invited Paper) [J]. International Journal of Computational Cognition, 2005, 3(2): 27-33

chio 算法,但是分类精度低于 Rocchio 算法。

表 12 各语料库在各个分类算法上的时间代价

语料库	分类算法	
	Knn	Rocchio
ChnSentiCorp-Book-del-4000	59	83
ChnSentiCorp-NB-del-4000	60	83

以上的实验结果表明,在选择特征提取、特征向量加权和情感分类算法时,需要根据用户的需求和数据集特征来进行算法组合,在分类精度和时间代价之间进行权衡。数据量大应该选择时间代价小的算法;如果数据量小,就选择精度高的算法,因为数据越小,各种算法的时间代价的差异越不明显。

结束语 云计算这一新型计算模型的兴起,为解决海量数据情感分类任务提供了可能性。本文利用 Hadoop 的 MapReduce 和 HDFS 技术,针对情感分类任务中特征提取、特征向量加权和分类等 3 个子任务,实现了 3 个子任务中涉及算法的 MapReduce 化。在情感语料数据集上进行了实验,针对每个子任务,选择不同的并行化算法进行组合,对每种组合下的情感分类精度及其每种算法的时间开销进行了对比分析,验证了本文实现算法的有效性,同时为用户选择不同算法实现情感分类任务提供了有价值的参考信息。

在本文工作的基础上,我们下一步将研究云平台下智能化的数据挖掘服务技术,根据用户的需求,考虑同类型不同算法的各种特征,如性能、精度等,自动化、智能地实现云平台下数据挖掘算法的组合,为用户提供智慧数据挖掘服务。

参考文献

[1] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification Using Machine Learning Techniques[C]// Proceedings of the EMNLP'02. 2002;79-86

[2] Pang B, Lee L. A Sentimental Education; Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts[C]// Proceeding of ACL. 2004;271-278

[3] Dave K, Lawrence S, Pennock D. Mining the peanut gallery: opinion extraction and semantic classification of product reviews [C]// Proceedings of WWW2003. 2003

[4] Mullen T, Collier N. Sentiment analysis using support vector machines with diverse information sources[C]// Proceedings of EMNLP'2004. 2004

[5] Li J, Sun M. Experimental Study on Sentiment Classification of Chinese Review using Machine Learning Techniques[C]// Proceedings of IEEE NLP-KE'2007. 2007

[6] Zhai Zhong-wu, Xu Hua, Li Juu, et al. Sentiment Classification for Chinese Reviews Based on Key Substring Features[C]// Proceedings of Natural Language Processing and Knowledge Engineering. 2009;24-27

[7] Devitt A, Ahmad K. Sentiment polarity identification in financial news: a cohesion-based approach[C]// Proceedings of ACL. 2007;984-991

[8] Shein K P P, Nyunt T T S. Sentiment classification based on Ontology and SVM Classifier[C]// Proceedings of ICCSN. 2010;169-172

[9] Zhai Zhong-wu, Xu Hua, Kang Ba-da, et al. Exploiting effective features for Chinese sentiment classification[J]. Expert Syst. Appl., 2011, 38(8):9139-9146

[10] Jeffrey D, Sanjay G. MapReduce: simplified data processing on large clusters[J]. Commun. ACM, 2008, 51(1):107-113

[11] Rocchio J. Relevance Feedback in Information Retrieval[M]. The SMART Retrieval System; Experiments in Automatic Document Processing, Chapter 14, Prentice-Hall, 1971;313-323

(上接第 205 页)

- [2] Papageorgiou Elpiniki I, Wojciech F. Application of Evolutionary Fuzzy Cognitive Maps for Prediction of Pulmonary Infections [J]. IEEE Transactions on Information Technology in Biomedicine, 2012, 16(1):143-149
- [3] Michael G. Fuzzy Cognitive Strategic Maps in Business Process Performance Measurement[J]. Expert Systems with Applications, 2013, 40(1):1-14
- [4] Kontogianni Areti D, Papageorgiou Elpiniki I, Christos T. How do you perceive environmental change? Fuzzy Cognitive Mapping informing stakeholder analysis for environmental policy making and non-market valuation[J]. Applied Soft Computing, 2012, 12(12):3725-3735
- [5] Kumar M V, Laurens B, Papageorgiou Elpiniki I, et al. Fuzzy cognitive maps and cellular automata: An evolutionary approach for social systems modeling[J]. Applied Soft Computing, 2012, 12(12):3771-3784
- [6] Peng Zhen, Yang Bing-ru, Wu Li-feng. The Construction of Fuzzy Cognitive Map Classification Model based on Relational Database[J]. International Journal of Digital Content Technology and its Applications, 2011, 5(5):33-41
- [7] 张桂云, 马希荣, 杨炳儒. 复杂模糊认知图的分解研究[J]. 计算机科学, 2007, 34(4):129-132
- [8] Peng Zhen, Wu Li-feng. Two-level Fuzzy Cognitive Map Mining for Text Categorization[J]. International Journal of Digital Content Technology and its Applications, 2012, 6(2):296-302
- [9] Lin Chun-mei. Combination study of multi fuzzy cognitive map [A]// Knowledge-Based Intelligent Information and Engineering Systems, 2008 [C]. Zagreb, 2008;332-339
- [10] Wojciech S, Lukasz K, Witold P. A Divide and Conquer Method for Learning Large Fuzzy Cognitive Maps[J]. Fuzzy Sets and Systems, 2010, 161:2515-2532
- [11] 骆祥峰. 认知图理论及其在图像分析与理解中的应用[D]. 合肥:合肥工业大学, 2003
- [12] 胡可云, 陆玉昌, 石纯一. 基于概念格的分类与关联规则的集成挖掘方法[J]. 软件学报, 2000, 11(11):1478-1484
- [13] Papageorgiou E I, Stylios C D, Groumpos P P. Active Hebbian Learning Algorithm to Train Fuzzy Cognitive Maps[J]. International Journal of Approximate Reasoning, 2004, 37(3):219-249
- [14] Papageorgiou E I, Stylios C D. Fuzzy Cognitive Map Learning based on Nonlinear Hebbian Rule[A]// Artificial Intelligence, 2003[C]. Australia, 2003;256-268
- [15] Sudjianto A, Hassoun M H. Statistical Basis of Nonlinear Hebbian Learning and Application to Clustering[J]. Neural Networks, 1995, 8(5):707-715
- [16] Wojciech S, Lukasz K, Witold P, et al. Genetic Learning of Fuzzy Cognitive Maps[J]. Fuzzy Sets and Systems, 2005, 153(3):371-401
- [17] Peng Zhen, Yang Bing-ru. Research on Balance of Fuzzy Cognitive Map[J]. Scientific Research and Essays, 2011, 6(31):6444-6447