

对基于知识发现的神经网络集成方法的研究^{*}

王 泳 邢红杰

(中国科学院自动化研究所模式识别国家重点实验室 北京 100080)¹

摘 要 该文对应用知识发现技术训练神经网络集成的方法进行了研究,提出了以并行操作的方式结合归纳学习所获取的知识和演绎学习所获取的知识的神经网络集成模型 KBNNE(Knowledge-based Neural Network Ensembles)。实验表明,通过调节所获取知识的权重因子,新模型可以有效提高网络集成的性能。

关键词 知识发现,神经网络集成,归纳学习,演绎学习

A Research on the Knowledge-based Neural Network Ensembles

WANG Yong XING Hong-Jie

(National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing 100080)¹

Abstract This article explores the utility of knowledge discovery in Neural Network Ensembles. A novel neural network ensemble model KBNNE(Knowledge-Based Neural Network Ensembles) integrating KDD(Knowledge Discovery in Database) techniques and neural network ensemble algorithms by parallel operations is proposed. By balancing the relative importance of knowledge learned by induction and deduction, the new model improves the quality of neural network ensembles and has been applied successfully to actual modeling problems.

Keywords Knowledge discovery, Neural network ensembles, Induction, Deduction

1 引言

神经网络的学习过程就是对周围环境的模拟过程,数据作为一种对周围环境的抽象描述(尽管可能有误差),在学习过程能够中发挥着不可替代的作用^[1]。虽然根据“没有免费的午餐定理”(No Free Lunch, NFL)^[2],在没有任何先验知识的情况下,只根据训练数据所获得的学习模型是不可靠的。但通过使用知识发现技术(Knowledge Discovery)^[3],完全可以从原始数据集中获得新的有关周围环境的信息,即实现了数据变换。根据更新后的训练数据所学习得到的模型,其可靠性完全可以得到保证。从这个角度上说,数据在知识发现技术和神经网络训练算法之间架起了沟通的桥梁。本文重点研究了在训练神经网络集成中应用知识发现技术的方法。

文章第 2 节简述了知识发现、数据变换和神经网络之间的关系;第 3 节提出了一种将知识发现技术与神经网络训练算法相结合的知识发现神经网络集成模型 KBNNE(Knowledge-Based Neural Network Ensemble);第 4 节是模拟实验;最后对文章内容进行了总结。

2 知识发现

知识发现是指数据库知识发现(Knowledge Discovery in Database, KDD),是一个从数据中获取有效的、新颖的、具有潜在使用价值的且最终可被理解的模式的过程^[3]。传统的数据分析方法(例如,统计方法)只能获得数据的表层信息,而 KDD 的目的则是寻找和观察数据之间的某种规律,增强数据的“可解释性”^[4],且在此过程中往往就包含了对原始数据的数据变换。从知识发现的主要步骤看^[5],在实施数据

挖掘(Data Mining)之前需要对原始数据进行整理(Data Cleaning),这其中包括对冗余数据的删除(Duplicate Elimination)、数据一致性分析(Consistency)、去除数据噪声(Noise Removal)、对数据中缺失信息的修复(Data Restoring);而在数据挖掘之后对知识进行描述(interpret),也就是采用语言、图形、可视化等手段对所获得的模式进行解释(规则提取),其中也包含着对数据的重新认识(如表 1)。

虽然传统上 KDD 是面向数据库等大型数据集合的^[6],但其数据挖掘方法大多是基于决策树、统计学习、机器学习、神经网络等^[7],因此 KDD 对于广义上的任何数据集合都是适用的。通过对数据的变换,一方面,神经网络方法成为知识发现的有力工具,另一方面,知识发现所获得的信息也有助于神经网络的训练。近年来对二者之间关系的研究日益受到人们的重视^[8]。

3 KBNNE

应用神经网络进行知识发现是一个归纳学习的过程,而将知识发现所获得的知识应用到神经网络的训练是一个演绎学习的过程,在人工智能领域已有学者就如何将归纳学习和演绎学习相结合的问题进行了研究^[25],而所获得的非常重要的结论就是将归纳学习和演绎学习以分步操作的方式进行结合的 EBNN(Explanation-Based Neural Network)^[26]学习算法。它将数据集 $D = \{(\bar{x}_i, y_i)\}_{i=1}^N$ (其中 $\bar{x}_i \in R^n, y_i \in R$) 划分为支持数据集(Support Set) $D_S = \{(\bar{x}_S, y_S)\}_{i=1}^p$ 和训练数据集(Training Set) $D_T = \{(\bar{x}_T, y_T)\}_{i=1}^q$,其中 $D_S \cap D_T = \emptyset$, $D_S \cup D_T = D$,通过神经网络对支持数据集中的数据进行训练学习,并从中归纳学习出一些通常用 IF...THEN 表示的域知识(Domain Knowledge),然后将所获得的域知识当作结构

^{*} 本文研究得到国家自然科学基金资助(#60275025, #60121302)。王 泳 博士研究生,主要研究方向为人工智能、模式识别、知识发现、数据挖掘。邢红杰 博士研究生,主要研究方向为智能系统、模式识别、数据挖掘。

约束结合训练数据集来训练新的神经网络,从而以串行操作的方式实现了归纳学习与演绎学习的结合(如图 1)。

表 1 应用 KDD 技术进行数据变换的方法

数据变换的步骤	具体方法说明	主要参考文献
删除冗余数据	删除重复数据和重复属性 (多出现于多源数据中)	Lee 和 Hsu [9]
数据一致性分析	调和数据内部(Self-Consistency)以及数据和参照数据(Relative-Consistency)之间的矛盾	Alimonte 和 Zibordi [10]
	调和数据和算法之间的矛盾 (Deterministic Learning, Active Learning)	Cervellera 和 Muselli [11]
去除数据噪声	提取数据主要特征(降维)	Duda 等[2] Tenenbaum 等[12]
恢复损失特征	(PCA, NLCA, ICA, MDS, Isomap, LLE)	Roweis 和 Saul [13]
	恢复数据原有特征(维数不变)	Arabie 等[14]
	(聚类, 层次聚类, 约束聚类)	Kaufman 和 Rousseeuw [15]
增加数据额外特征	考虑数据之间的相关性或时序性	Cao 和 Tay [16]
	给数据添加额外的信息特征(升维)	
去除数据噪声	对单点数据损失使用各种滤波	Awate 和 Whitaker [17]
恢复损失数据	方法或相邻点平均法进行恢复	Tholudur 等[18]
	对局部区域数据损失使用变换	
	域修复法或空域修复法恢复	Sun 等[19]
忽略无法恢复数据	降低损失数据对模型的影响	Little 和 Rubin [20]
	(最大似然估计, 贝叶斯估计)	
规则提取	以语言的形式描述规则	Zhou 和 Sun [21]
	以图形的形式描述规则	Jordan 和 Seinowski [22]
	以数学表达的形式(假设检验)描述规则	Garson [23]
	以可视化技术描述规则	Keller 等[24]

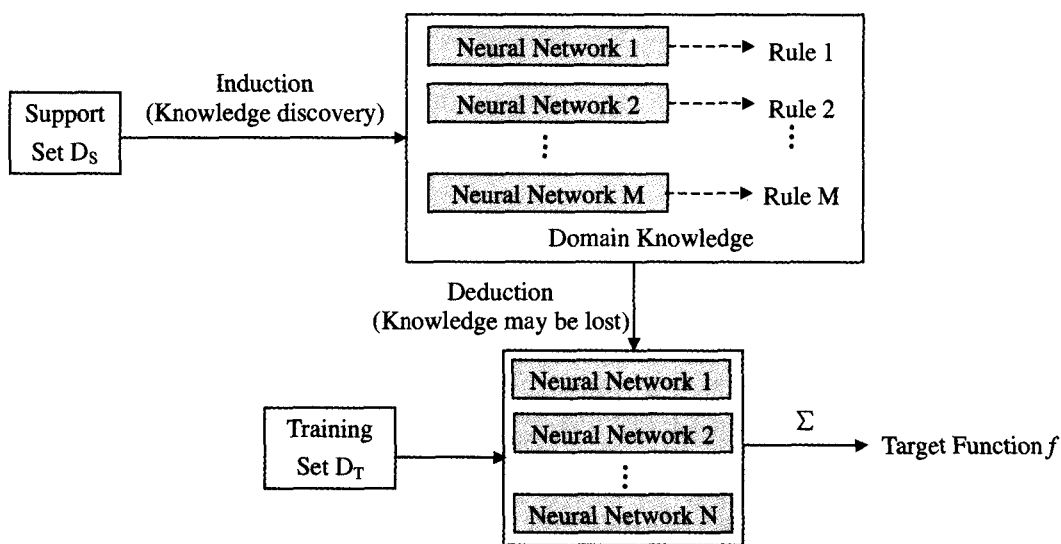


图 1 EBNN-以串行操作的方式结合归纳学习和演绎学习

如果用 U 代表全部的域知识, $L_S = \{L_{S1}, L_{S2}, \dots, L_{SM}\}$ 代表对支持数据集进行训练的神经网络所代表的域知识, $R_S = \{R_{S1}, R_{S2}, \dots, R_{SM}\}$ 代表归纳得出的学习规则, $L_T = \{L_{T1}, L_{T2}, \dots, L_{TN}\}$ 代表由训练数据集演绎得出的域知识, 则有以下式成立:

$$R_S \subseteq L_S \subseteq U \Leftrightarrow R_{S1} \cup R_{S2} \cup \dots \cup R_{SM} \subseteq L_{S1} \cup L_{S2} \cup \dots \cup L_{SM} \subseteq U \quad (1)$$

$$R_S \subseteq L_T \subseteq U \Leftrightarrow R_{S1} \cup R_{S2} \cup \dots \cup R_{SM} \subseteq L_{T1} \cup L_{T2} \cup \dots \cup L_{TN} \subseteq U \quad (2)$$

最终, 通过对每一个演绎规则施加重 w_i , 构造网络集成对目标函数进行估计(式 3)

$$f \approx \sum_{i=1}^N w_i L_{Ti}, \left(\sum_{i=1}^N w_i = 1, w_i \geq 0 \right) \quad (3)$$

虽然 R_S 来自于对 L_S 的归纳, L_T 是对 R_S 的演绎, 但在串行操作的情况下 L_S 和 L_T 并没有直接的联系, 而且 L_S 和 L_T 都是以神经网络的形式代表域知识, 而 R_S 是以符号语言的形式代表域知识, 因此在知识描述的转化过程中很难避免知识的遗漏 (Knowledge Leakage), 即:

$$L_S \cup L_T \neq L_T \quad (4)$$

为了克服由于知识描述方式不同所产生的知识遗漏, KBNNE 将以 L_S 表示的神经网络和以 L_T 表示的神经网络进行了集成(如图 2)。在集成过程中, KBNNE 根据归纳学习所获得的知识动态地对训练数据集进行调整, 同时通过权重因子 μ_S 和 μ_T 来调节所获取知识对最终结果的影响程度, 并将归纳学习和演绎学习以并行的方式进行了结合(式 5)。当 μ_S 取值为 0, μ_T 取值为 1 时, 式 (5) 就退化为式 (3), 因此 KBNNE 是对 EBNN 的扩展。

$$f \approx \mu_S L_S + \mu_T L_T = \mu_S \sum_{i=1}^M w_{Si} L_{Si} + \mu_T \sum_{i=1}^N w_{Ti} L_{Ti} \quad (5)$$

$$\sum_{i=1}^M w_{Si} = 1, w_{Si} \geq 0; \sum_{i=1}^N w_{Ti} = 1, w_{Ti} \geq 0$$

$$\mu_S + \mu_T = 1, \mu_S \geq 0, \mu_T \geq 0$$

根据已有研究, 在网络集成中使用加权平均有可能导致过匹配^[27], 因此在 KBNNE 中使用简单平均法构造网络集成。同时, 考虑到演绎学习所获得的知识比归纳学习所获得的知识更能影响最终学习的结果, 因此 μ_T 的取值一般要比 μ_S 大(式 (6))。

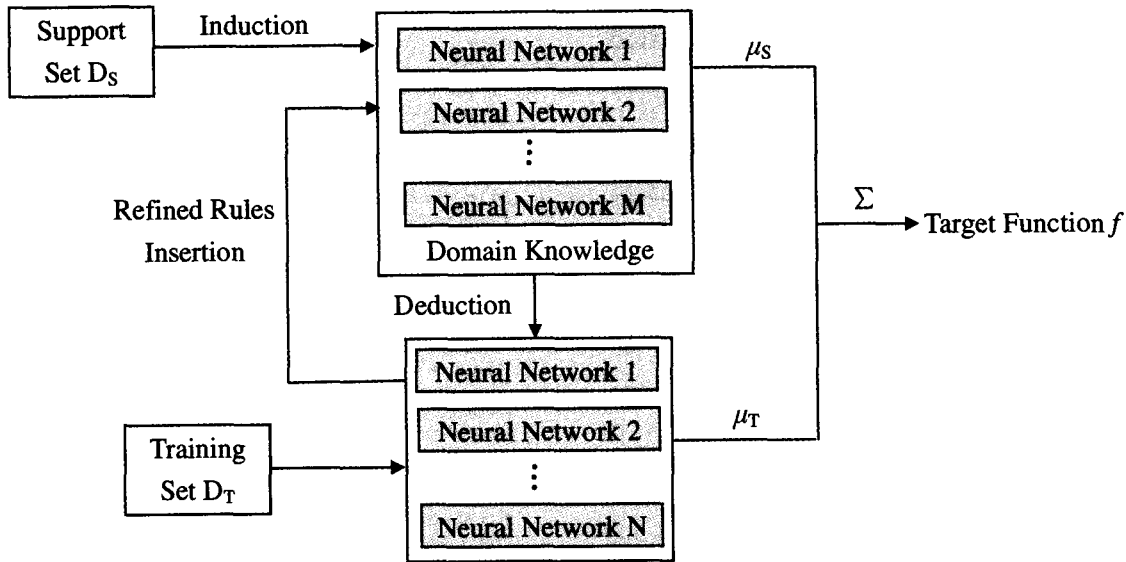


图2 KBNNE-以并行操作方式结合归纳学习和演绎学习

$$f \approx m_S L_S + m_T L_T = \frac{m_S \sum_{i=1}^M L_{S_i} + m_T \sum_{i=1}^N L_{T_i}}{M+N}$$
$$m_S + m_T = 1, m_T > m_S \geq 0 \quad (6)$$

4 实验结果

实验全部是在 Scilab 科学计算软件环境下完成(<http://scilabsoft.inria.fr>)。KBNNE 模型由一条归纳准则(一个 SVM 分类器)和两条演绎准则(两个单隐层前馈神经网络)组成。SVM 方法和基于 Shepard 插值的核函数方法^[28]中都使用 RBF 核函数,神经网络集成由 3 个前馈神经网络组成。为保证实验结果的公平,实验中的模型都取相同的参数值(例如,RBF 核函数中的参数,神经网络中都使用 Sigmoid 激活函数,隐层节点数都是 20 等)。

实验 1 本实验数据来自人造数据。对测试神经网络计算能力的标准异或问题^[1]进行扩展,随机产生先验概率分布分别为 0.5 的两类分类数据:

- 第一类样本由中心点分别为 $(-0.5, 0.3)$ 和 $(0.3, 0.6)$ 的二维高斯分布函数产生,标准偏差 (standard deviation) 都取值 0.05,其中由第一个中心点产生的数据占全部第一类数据的 10/11,由第二个中心点产生的数据占全部第一类数据的 1/11(如图 3 中 ∇ 点所示)。

- 第二类样本由中心点分别为 $(0.5, 0.6)$ 和 $(-0.3, 0.3)$ 的二维高斯分布函数产生,标准偏差都取值 0.03,其中由第一个中心点产生的数据占全部第二类数据的 10/11,由第二个中心点产生的数据占全部第二类数据的 1/11(如图 3 中 $\circ$ 点所示)。

训练数据集由 110 个样本组成,测试数据集由 1100 个样本组成。我们将 KBNNE 方法与 SVM 方法和神经网络集成方法进行了对比测试,图 3 显示了不同学习模型产生的不同分类界面(RBF 核中参数 $g=1, C=100$),表 2 列出了不同学习模型在不同模型参数下的分类误差,实验结果表明 KBNNE 模型可以有效利用知识,降低分类误差。

实验 2 本实验数据来自 Increment SVM 软件包中所带两类分类基准测试数据 nonlineardata100(<http://bach.ece.jhu.edu/pub/gert/svm/incremental>)。我们分别使用不同的

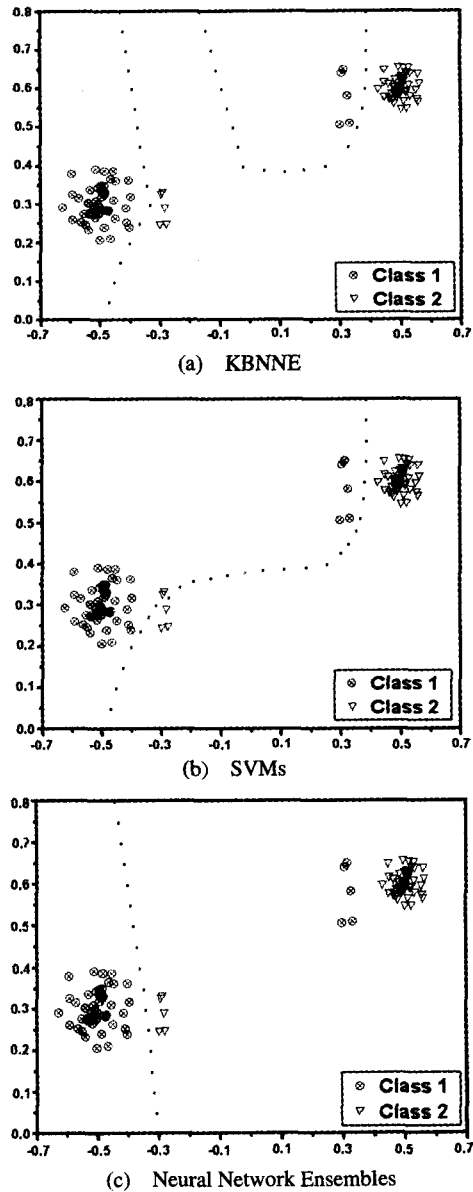


图3 不同学习模型所产生的分类界面(RBF 核参数 $g=1, C=100$)

表2 实验1中不同学习模型在不同模型参数下的分类误差

学习模型	模型参数	训练误差	测试误差
KBNNE	$g=1, C=100$	0.0%	1.0%
	隐层节点数 20		
	$g=0.5, C=100$	2.73%	4.18%
SVMs	$g=1, C=100$	1.82%	1.91%
	隐层节点数 20		
	$g=0.5, C=100$	5.45%	5.91%
Neural Network Ensembles	隐层节点数 20	10.91%	12.45%

训练样本个数(20,30,40,50)来检测模型的泛化能力。测试样本个数都是50。我们将KBNNE与基于Shepard插值的核函数方法、SVM方法和神经网络集成方法进行了对比测试。实验结果见图4。

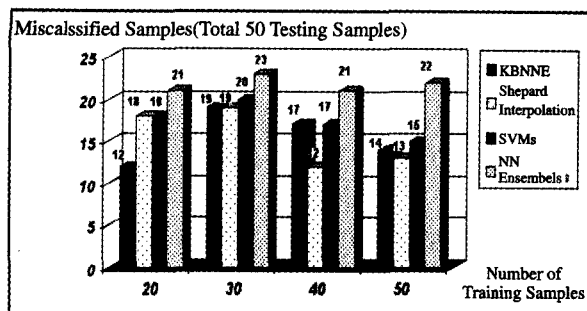


图4 KBNNE模型与其它学习算法的对比

可以看出,基于知识发现的神经网络集成模型KBNNE的泛化性能显得很稳定。尽管有时候在所有方法中它的判别准确率不是最高的,但也远远高于一般的神经网络集成的方法,并且与最佳值比较接近。

结论 本文从知识发现的角度出发,以构造神经网络集成的方式将知识发现技术与神经网络训练算法进行结合,实现了归纳学习和演绎学习的有效互补。与传统的EBNN模型不同,本文提出的KBNNE模型以并行操作的方式建立了知识的描述方式和网络训练之间内部的联系,通过对训练数据的动态变换,不仅增强了模型的适用范围,而且对于解决神经网络以何种方式应用知识,如何理解网络计算的“可靠性”、“透明度”等有关神经网络本质的问题都是有益的探索。

参考文献

- Haykin S. Neural Networks: A Comprehensive Foundation (2nd edition). Prentice-Hall, New Jersey, 1999
- Duda R O, Hart P E, Stork D. Pattern Classification (2nd edition). New York: John Wiley, 2001
- Fayyad U M, Shapiro G P, Smyth P, Uthurusamy R. Advances

- in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996
- 郭萌,王珏. 数据挖掘与数据库知识发现:综述. 模式识别与人工智能, 1998, 11(3): 292~299
- Fayyad U M, Shapiro G P, Smyth P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, 1996, 39(11): 27~34
- Fayyad U M. Data Mining and Knowledge Discovery: Making Sense Out of Data. IEEE Expert, 1996, 11(5): 20~25
- Mitra S, Pal S K, Mitra P. Data Mining in Soft Computing Framework: A Survey. IEEE Trans. on Neural Networks, 2002, 13(1): 3~14
- 员巧云,程刚. 近年来我国数据挖掘研究综述. 情报学报, 2005, 24(2): 250~256
- Lee M L, Hsu W. Improving data quality: eliminating dupes & I-D-ing those spurious links. IEEE Potentials, 2005, 24(2): 35~38
- Alimonte D D, Zibordi G. Statistical Assessment of Radiometric Measurements From Autonomous Systems. IEEE Trans. on Geoscience and Remote Sensing, 2006, 44(3): 719~728
- Cervellera C, Muselli M. Deterministic Design for Neural Network Learning: An Approach Based on Discrepancy. IEEE Trans. on Neural Networks, 2004, 15(3): 533~544
- Tenenbaum J B, Silva V, Langford J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, 2000, 290(5500): 2319~2323
- Roweis S T, Saul L K. Nonlinear Dimensionality Reduction by Local Linear Embedding. Science, 2000, 290(5500): 2323~2326
- Arabie P, Hubert L J, Soete G D. Clustering and Classification. World Scientific, 1996
- Kaufman L, Rousseeuw P J. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series In Probability And Mathematical Statistics, New York: John Wiley, 2005
- Cao L J, Tay F E H. Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. IEEE Trans. on Neural Networks, 2003, 14(6): 1506~1518
- Awate S P, Whitaker R T. Unsupervised, Information-Theoretic, Adaptive Image Filtering for Image Restoration. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006, 28(3): 364~376
- Tholudur A, Ramirez W F, McMillan J D. Interpolated Parameter Functions for Neural Network Models. Computers and Chemical Engineering, 2000, 24: 2545~2553
- 孙晓军,丁群,朱勇. 乘窗信号谱分解及其数据修复. 电子学报, 2005, 33(3): 538~540
- Little R J A, Rubin D B. Statistical Analysis with Missing Data (2nd edition). Wiley Series In Probability And Statistics, New York: John Wiley, 2002
- 周志华,陈世福. 神经网络规则抽取. 计算机研究与发展, 2002, 39(4): 398~405
- Jordan M I, Sejnowski T J. Graphical Models: Foundations of Neural Computation. MIT Press, 2001
- Garson G D. Interpreting neural-network connection weights. AI Expert, 1991, 6(4): 47~51
- Keller T, Gerjets P, Scheiter K, et al. Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. Computers in Human Behavior, 2006, 22: 43~65
- Bergadano F, Giordana A, Saitta L. Machine Learning: An Integrated Framework and its Application. Ellis Horwood Series in Artificial Intelligence, Ellis Horwood, 1991
- Thrun S. Explanation-Based Neural Network Learning: A Life-long Learning Approach. Kluwer Academic, Boston, 1996
- Opitz D, Shavlik J. Actively searching for an effective neural network ensemble. Connection Science, 1996, 8(3-4): 337~353
- 吴涛,贺函根,贺明科. 基于插值的核函数构造. 计算机学报, 2003, 26(8): 990~996

(上接第171页)

的算法。文中的分析和仿真实验的结果表明,本文提出的MIOFN运行是可靠的,在工程应用中是有效、可行的。

参考文献

- Castillo E. Functional networks[J]. Neural Processing Letters, 1998, 7: 151~159
- Noelia S M, Oscar F R, Amparo A B. Self-Organizing Maps and Functional Networks for Local Dynamic Modeling [C]. In: European Symposium on Artificial Neural Networks Bruges (Belgium). April 2003, 39~44
- Castillo E, Gutiérrez J M. Some Applications of Functional Networks in Statistics and Engineering [J]. Technometrics, 2001, 43(1): 10~24

- Castillo E, Cobo A, Gutiérrez J M. Working with Differential, Functional and Difference Equations Using Functional Networks [J]. Applied Mathematical Modeling, 1999, 23: 89~107
- 周永权,焦李成. 层次泛函网络整体学习算法. 计算机学报[J], 2005, 28(8): 1277~1286
- Castillo E, Ruiz-Cobo R. Functional Equations in Science and Engineering [M]. New York: Marcel Dekker, 1992
- Castillo E. Functional networks [J]. Neural Processing Letters, 1998, 7: 151~159
- Castillo E, Gutiérrez J M. A Minimax Method for Learning Functional Networks [J]. Neural Processing Letters, 2000, 11: 39~49
- Zienkiewicz O C. Flow Formulation for Numerical Solution of Forming Processes [C]. Numerical Analysis of Forming Processes. New York: Wiley, 1984. 1~44