

基于 RBF 神经网络的医院科技影响力评价模型构建^{*}

安新颖 代涛 钱庆

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 在文献调研的基础上基于德尔菲法建立医院科技影响力评价指标体系,根据专家经验和知识对 60 家医院的科技影响力进行排名。采用径向基函数神经网络方法,将 50 个医院作为训练样本,10 个医院作为测试样本,构建出一套基于径向基函数的医院科技影响力的评价模型,实验结果表明该评价模型较好地拟合了专家的思维,可以使医院科技影响力评价真实有效。

〔关键词〕 医院科技影响力;径向基函数神经网络;评价

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2016.03.002

Building a Sci-tech Influence Evaluation Model for Hospitals Based on RBF Neural Network AN Xin-ying, DAI Tao, QIAN Qing, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 Based on literature research, the paper builds a sci-tech influence evaluation index system for hospitals with the Delphi method and ranks the sci-tech influence of 60 hospitals in terms of expertise. By using radial basis function neural network and taking 50 hospitals as training samples and 10 hospitals as test samples, it builds a sci-tech influence evaluation model for the hospitals based on the radial basis function. As indicated by experimental results, this evaluation model fits thoughts of experts well and evaluates the sci-tech influence of hospitals truly and effectively.

〔Keywords〕 Hospital's sci-tech influence; Radial basis function neural network; Evaluation

1 医院科技影响力

随着对科技资源使用效益的重视及对科研考核的需要,越来越多的国内外科研管理人员开始应用

量化的评估体系,在科研人员、团队、科研机构等不同层次上开展科技影响力评价。医院是一类特殊的科研机构,其科技水平也在某种程度上反映其整体技术水平,是医院可持续发展的重要保证。

医院科技影响力是指医院经过知识和技能的积累,使其在竞争环境中能够长时间取得主动的能力,这是医院对能够确保竞争优势的组织资本和社会资本的有机结合与发挥,反映了医院科技活动影响的范围和深度。目前,一些机构已经开展了医院评价相关工作。国家卫生和计划生育委员会制订的“综合医院评价标准”^[1]中,应用近 5 年国内论文数 ISSN、国内论文数及被引用次数、SCI 收录论文数,

〔收稿日期〕 2016-02-28

〔作者简介〕 安新颖,博士,副研究员,发表论文 40 余篇;通讯作者:钱庆。

〔作者简介〕 中国医学科学院医学信息研究所中央公益性基本科研业务课题“中国医院科技影响力评价的关键问题研究”(项目编号:15R011)。

承担与完成国家、省级科研课题数, 获得国家、省级科研基金额度等方面来评价综合医院的科技水平。复旦大学医院管理研究所开展的“中国最佳医院及最佳专科声誉排行榜”^[2], 以学科声誉为主, 兼顾学术, 医院学科声誉权重 80%, 学术科研权重 20%。中国科学技术信息研究所对医院科技论文被引用篇数、被引用次数, 以及 SCI 和 Medline 论文数的增长趋势进行了统计并对近 10 年国际论文累计被引用篇数较多的前 20 家医院和拥有表现不俗的论文数量较多的医疗机构进行了排名。沈林等人对杭州市级医疗卫生机构科研竞争力开展了评价研究^[3]。在医院科技影响力评价的相关研究中较多采用层次分析法或德尔菲法等主观赋权法来确定指标的权重, 客观性较差^[4]。此外, 科技影响力评价指标与目标之间不完全是线性函数关系, 通常所采用的线性加权方法不一定适用。随着人工智能的迅速发展, 人工神经网络具有自学习、自组织能力, 可较好的处理非线性关系, 为开展复杂的医院科技影响力评价创造了条件。

2 径向基函数神经网络

2.1 径向基函数神经网络简介

人工神经网络是 20 世纪 80 年代以后迅速发展起来的一种网络技术, 它是由大量简单的神经元相互连接而成的自适应非线性动态系统^[1]。在人们提出的几十种神经网络模型中, 用的最多的是反向传播 (Back Propagation, BP) 网络、Hopfield 网络、自适应共振理论 (Adaptive Resonate Theory, ART) 网络以及径向基函数 (Radial Basis Function, RBF) 网络。BP 网络由于其网络收敛速度较慢、且容易陷入局部最小值, 导致其应用受到限制。径向基函数神经网络 (Radial Basis Function, RBF) 是 1985 年 Powell M. J. D 提出的一种以函数逼近理论为基础构造、具有单个隐藏层的 3 层前馈型神经网络^[1], 包括一个输入层、一个具有径向基函数神经元的隐藏层和一个具有线性神经元的输出层, 见图 1。

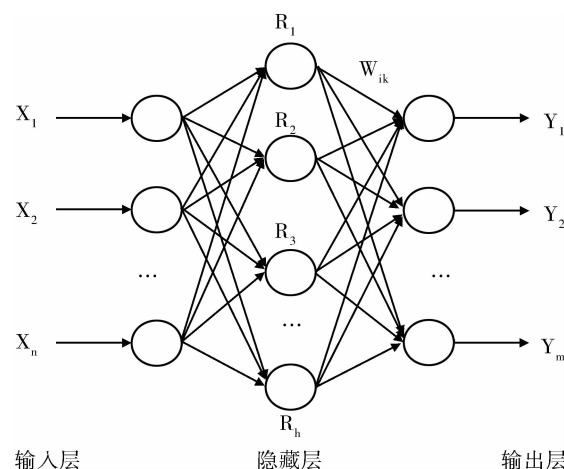


图 1 RBF 神经网络结构

隐藏层的作用是将输入矢量从低维空间映射到多维空间, 一般采用高斯函数作为其节点的基函数, 其形式见公式 1。 $R_i(X)$ 为第 i 个高斯基函数的宽度, x 为输入向量, C_i 为第 i 个隐节点的中心, σ_i 为宽度参数, m 为隐节点个数。RBF 网络的学习首先是初始化, 根据所有输入样本确定隐藏层各节点高斯基函数的中心值和宽度参数。然后利用最小二乘法, 求出输出单元权值。RBF 神经网络通过适当地调整隐藏层节点数、中心值以及隐藏层到输出层的连接权值, 输出值见公式 2, W_{ij} 为隐藏层 i 神经元和输出层 j 神经元之间的权值, a_j 为输出层的阈值^[7]。

$$R_i(X) = \exp\left(-\frac{\|x - C_i\|^2}{2\sigma_i^2}\right) (i = 1, 2, \dots, m) \quad (1)$$

$$y_j = \sum_i W_{ij} R_i(X) + a_j \quad (2)$$

RBF 为局部逼近网络, 具有学习速度快的优点, 因此 RBF 在逼近能力、学习速度以及收敛速度方面均优于 BP 神经网络。

2.2 径向基函数神经网络应用

国外 RBF 神经网络在疾病分类^[8]、面部识别^[9-10]以及模式识别领域得到了较好的应用, 国内经过几年的发展除了在上述领域广泛应用 RBF 神经网络外, 在环境预测、信用评估、投资风险评估以及信息素养评价等领域也开展了探索研究。Wei - zhen Lu 应用主成分分析和 RBF 对香港的污染情况

进行了预测^[11]。Tuba KIYAN 和 Tulay YILDIRIM 利用 RBF 神经网络进行乳腺癌的诊断^[12]。A. Najah 和 A. El – Shafie 等人分别利用线性回归模型和 RBF 神经网络进行了地表水污染的预测，试验证明 RBF 神经网络在揭示水质量的相关参数时更为准确^[13]。Hua Jiang 将神经网络应用于高技术项目投资风险评估，研究结果显示 RBF 比 BP 神经网络在进行高技术项目风险评估更加有效^[14]。王辅之和罗爱静等人基于 AHP – RBF 神经网络开展了居民健康信息素养评价模型研究^[15]，张筠莉和杨祯山利用灰色 RBF 神经网络开展现代医院门诊服务能力预测^[16]。

本文在文献调查和德尔菲法的基础上，建立医院科技影响力评价指标体系，经专家咨询，对 60 家医院进行了评价。为模拟专家做出评价的过程，本研究以该评价结果中 50 家医院数据作为训练样本，采用 RBF 神经网络模拟专家评价模式，利用 10 家医院数据进行模拟，以分析该神经网络的实际输出和专家评价结果之间的误差。该研究可以建立一套较为科学、有效的医院科技影响力评价模型，降低专家在评价过程中的主观性，为开展科研机构评价提供了新思路和新方法。

3 基于RBF神经网络的医院科技影响力评价模型构建

3.1 指标及权重设计

本文在大量阅读相关研究文献的基础上，结合医药卫生领域相关专家的意见，基于科学性、全面性以及可获得性原则，构建了科技投入、科技产出以及科技影响 3 个 1 级指标，9 个 2 级指标和 21 个 3 级指标来全面反映医院科技影响力水平，见表 1。

表 1 医院科技影响力评价指标体系		
1 级指标	2 级指标	3 级指标
科技投入	科研人员	科研人员数量
	科研项目	国家级科研项目数量 部级科研项目数量 临床试验项目数量
	科研平台	国家级/部级重点实验室数量 国家临床医学研究中心数量 国家临床重点专科数量 药物临床试验机构专业数量

续表 1

科技产出	期刊论文	SCIE 收录论文数量 中文核心期刊论文量 SCIE 收录论文总被引频次 中文核心期刊论文总被引频次
		重要国际会议论文数量
		发明专利授权量
		国家标准和卫生行业标准数量
学术影响	杰出人才和团队	杰出人才数量 杰出团队数量
		重要学会任职人数 国际重要检索系统收录的国内
	学术任职	期刊任职人数 国家科技奖数量
		科技奖项 中华医学科技奖数量

3.2 RBF 神经网络设计

3.2.1 确定训练样本和测试样本 训练样本对 RBF 网络模型的质量有一定决定作用，在医院科技影响力评价中要结合已经确定的训练样本的评价结果进行综合分析。

3.2.2 标准化 量纲的不同会导致评价结果的偏差，因此在开展评价之前需要对各个指标进行无量纲化处理。无量纲化处理的方法有很多种，本文采用以下函数转换公式进行数据的归一化，其中 X 为实际测量数据，Y 为归一化后的数据，*MaxValue* 为实际测量数据的最大值，*MinValue* 为实际测量数据的最小值。经过处理后的各项指标数值在 [0, 1] 之间，方便开展 RBF 计算。

$$Y = \frac{X - MinValue}{MaxValue - MinValue}$$

(3)

3.2.3 建立输入输出向量 输入向量是训练样本的各个相关评价指标的集合，即输入向量 $X_{ij} = [x_{i1}, x_{i2}, \dots, x_{ij}]$ ，输出向量是训练样本的综合判断结果，即输出向量 $Y_i = [y_1, y_2, \dots, y_i]$ ，输入向量和输出向量的主要作用是通过神经网络对训练样本的评价指标进行自动学习；

3.2.4 设计 RBF 神经网络 采用 NEWRBE (P, T, SPREAD) 函数进行训练，P 为输入向量，T 为期望输出向量，SPREAD 为径向基函数的分布密度，缺省值为 1.0。通过调整 SPREAD 的值来改变网络预测精度，从而找出目标输出与期望输出最接近的网络模型。若 SPREAD 的值较大，则相应的函数拟

合就更平滑，但 SPREAD 的值不是越大越好，若其值太大则反而会出现过拟合的情况，因为过大的扩展系数是需要非常多的神经元来适应函数的快速变化的。

3.2.5 测试样本评价 利用神经网络仿真函数 sim (Net, PC) 对测试样本进行评价，Net 为构建好的神经网络对象，PC 为测试样本。利用均方误差 (Mean Square Error, MSE) 来比较不同 Spread 值下网络性能的优劣。

$$MSE = \frac{1}{n} \sum_i^n (observed - predicted)^2 \tag{4}$$

4 实证分析

根据评价指标体系，收集医院相关指标数据，数据来源包括：有关政府部门的统计数据资料；国内、外有关数据库；有关医院的网站；有关内部资料等。在论文数据方面，国外部分统计拟选择 SCIE，国内部分统计选择 SinoMed 中文生物医学文献数据库，下载数据并构建数据集。数据集建立后，还需要对原始数据进行了全面核查，对异常数据进行处理，以及进行数据清洗及数据融合。本文采用 Matlab7.1 和 Excel 作为分析工具，以北京、上

海、广州、成都等地的 60 家三甲医院作为分析对象，为了使建立的 RBF 网络预测误差最小，将其中的 50 家作为训练样本用于 RBF 神经网络学习，10 家作为测试样本。按照 RBF 神经网络设计步骤 (2) 对数据进行标准化处理。根据步骤 (3) 和 (4) 构建 RBF 神经网络，见图 2。SPREAD 为 1.0 时，输入层神经元为 21，隐藏层神经元为 50，输出层神经元为 1，IW {1, 1} 和 b {1} 分别表示初始权重和阈值，LW {2, 1} 和 b {2} 为经过运算得到的新权重和阈值。RBF 神经网络构建后，利用步骤 (5) 对测试样本的目标结果进行仿真。Y = [0.2626, 0.2585, 0.2583, 0.2573, 0.2563, 0.2562, 0.2491, 0.2464, 0.2405, 0.2397]。由于不同的 SPREAD 值对 RBF 神经网络有一定影响，因此本研究对不同的 SPREAD 值 (1 ~ 100) 情况下的测试样本的期望输出进行了仿真，期望输出和实际输出之间的误差，见表 2，其中 SPREAD = 35 时，MSE 值最小，其网络性能最佳。

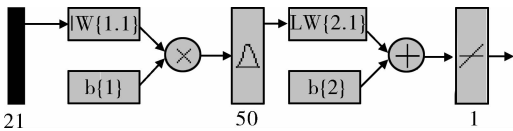


图2 建立的 RBF 神经网络模型

表2 期望输出和实际输出之间的误差

样本	期望输出	Spread 1.0	Spread 5.0	Spread 10	Spread 15	Spread 20	spread 25	Spread 30	spread 35
1	0.262 6	0.255 33	0.259 09	0.259 54	0.259 62	0.259 65	0.259 66	0.259 67	0.259 68
2	0.258 5	0.249 29	0.254 88	0.255 68	0.255 84	0.255 9	0.255 92	0.255 94	0.255 94
3	0.258 3	0.252 23	0.254 85	0.255 57	0.255 71	0.255 76	0.255 78	0.255 79	0.255 8
4	0.257 3	0.260 96	0.258 71	0.258 07	0.257 93	0.257 88	0.257 86	0.257 85	0.257 84
5	0.256 3	0.259 86	0.256 36	0.255 52	0.255 34	0.255 28	0.255 25	0.255 23	0.255 22
6	0.256 2	0.269 9	0.260 96	0.258 99	0.258 57	0.258 43	0.258 36	0.258 32	0.258 29
7	0.249 1	0.257 73	0.250 98	0.249 95	0.249 75	0.249 67	0.249 64	0.249 62	0.249 61
8	0.246 4	0.244 14	0.244 94	0.244 93	0.244 92	0.244 92	0.244 92	0.244 92	0.244 92
9	0.240 5	0.241 89	0.240 45	0.239 6	0.239 42	0.239 35	0.239 32	0.239 31	0.239 3
10	0.239 7	0.247 03	0.241 76	0.240 63	0.240 39	0.240 31	0.240 27	0.240 25	0.240 23
MSE	—	5.235 27E-05	7.188 84E-06	3.831 26E-06	3.385 49E-06	3.247 37E-06	3.193 14E-06	3.155 98E-06	3.132 75E-06

* * 由于篇幅原因，仅列出 8 个 spread 取值的神经网络仿真值。

5 结论

5.1 利用 RBF 神经网络评价医院科技影响力的优势

在医院科技影响力评价的方法中，最简单同时

也是被广泛使用的方法就是德尔菲法，根据专家对各个医院情况的了解进行综合评分。然而，虽然专家对各个医院的总体情况比较了解，但是对各个医院的所有指标的深刻理解仍存在主观上的差异。RBF 可以通过给定样本专家评价模式的学习，获取

专家评价经验、知识以及评价目标重要性的倾向, 然后对测试样本进行评价避免了专家个人主观认识受不同环境、不同时期的影响而产生波动, 做到评价结果的客观、科学。因此, RBF 是用定量的方法替代了专家主观经验的定性分析, 同时, 摆脱了评价时多次征询专家意见的繁琐过程, 也将分析人员从大量的计算中解脱出来, 使基于多维度指标计算的医院科技影响力评价更加科学、有效。

5.2 RBF 神经网络评价医院科技影响力的科学性

径向基函数神经网络的网络性能好坏, 与选择的训练样本有较为密切的关系。样本的规模和样本的质量对训练样本至关重要。本研究在大量影响医院科技影响力的相关理论和方法的基础上, 选择了 40 多个的指标集合, 但结合其科学性、合理性以及数据的可获得性, 根据专家的意见, 最终遴选出 21 个指标。为了确保训练样本的规模, 本文采用 50 组数据作为训练样本, 训练样本中包括了各类数据, 样本数据的获得全部采用定量化测定, 数据准确可靠, 提高了评价的客观性、准确性。然而, RBF 尚存在一些缺点, 由于其评价模型的隐含性导致其应用受到限制。它还不能提供解析表达式, 指标和目标之前的关系由于不是一种简单的线性关系, 因此权重还不能解释为一种回归系数, 也不能用来分析因果关系, 目前还不能从理论上解释 RBF 神经网络权值的意义, 因此, 要结合其他方法使其神经网络可读, 结果容易理解, 还要从神经元的个数、径向基函数的分布密度等参数角度进一步改进 RBF 神经网络。

参考文献

- 1 中华人民共和国卫生部. 综合医院评价标准 (修订稿) [EB/OL]. [2016-01-05]. <http://www.moh.gov.cn/publicfiles/business/cmsresources/mohbgt/cmsrsdocument/doc6535.pdf>.
- 2 复旦大学医院管理研究所. 2009 年度中国最佳医院及最佳专科声誉排行榜 [EB/OL]. [2016-01-05]. <http://www.fudanmed.com/institute/news222.aspx>.
- 3 沈林, 杨丽静, 王珍, 等. 杭州市级医疗卫生机构科研竞争力评价研究 [J]. 中华医学科研管理杂志, 2013,

- 26 (3): 162-165.
- 4 余春雨, 梁祥君. 进化神经网络法在科研项目评价中的应用 [J]. 中国科技论坛, 2007, (6): 107-109.
- 5 许月卿, 李双成, 蔡运龙. 基于 GIS 和人工神经网络的区域贫困化空间模拟分析 [J]. 地理科学进展, 2006, 25, (3): 79-85
- 6 Simon Haykin. Neural Networks; a Comprehensive Foundation [M]. Second edition. New Jersey Prentice Hall, 1998: 161-175.
- 7 李磊, 李剑, 马建华. RBF 神经网络在土壤重金属污染评价中的应用 [J]. 环境科学与技术, 2010, 33 (5): 191-195.
- 8 Sanjivani Bhande. Ranjana Raut. Intelligent Decision Support System for Parkinson Disease Using Softcomputing [J]. IOSR Journal of Electrical and Electronics Engineering, 2014, 17-22.
- 9 Meng JooEr, Shigion Wu, Juwei Lu, et al. Face Recognition with Radial basis Function (RBF) Neural Networks [J]. Applied soft computing, 2007, 7 (1): 58-70.
- 10 JamunaKanta Sing, Dipak Kumar Basu, mita Nasipari, et al. Face recognition Using Point Symmetry Distance-based RBF Network [J]. Applied Soft Computing, 2007, 7 (1): 58-70.
- 11 Wei Zhen Lu, Wen Jian Wang, Sui Huang Lan, et al. Potential Assessment of a Neural Network Model with PCA/RBF Approach for Forecasting Pollutant Trends in MongKok Urban air, Hong Kong [J]. Environmental Research, 2004, 96 (1): 79-87.
- 12 Tuba Kiyan, Tulay Yildirim. Breast Cancer Diagnosis Using Statistical Neural Networks [J]. Journal of Electrical & Electronics Engineering, 2004, 4 (2): 1149-1153.
- 13 A. Najah, A EI-Shafie, O. A. Karim, et al. Application of Artificial Neural Networks for Water Quality Prediction [J]. Neural Computing and Applications, 2013, 22 (suppl.): 187-201.
- 14 Hua Jiang. The Application of Artificial Neural Networks in Risk Assessment on High-tech Project Investment [C]. International Conference on Business Intelligence and Financial Engineering, 2009. BIFE 09. 24-26 July 2009, 17-20.
- 15 王辅之, 罗爱静, 孙伟伟, 等. 基于 AHP-RBF 神经网络的居民健康信息素养评价模型研究 [J]. 医学信息学杂志, 2013, 34 (7): 14-18.
- 16 张筠莉, 杨祯山. 现代医院门诊量的灰色 RBF 神经网络预测 [J]. 计算机工程与应用, 2010, 46 (29): 229-232.