

神经网络与 Logistic 回归在基础研究 成果综合评价中的应用比较

何钦成¹, 黄亚明¹, 何 苗²

(1. 中国医科大学 信息管理与信息系统(医学)系; 2. 中国医科大学 附属第一医院计算机室, 辽宁 沈阳 110001)

摘 要:应用现代数理统计、系统综合评价及科学计量学方法构建对基础性研究成果的综合评价系统。在以科学计量学和数理统计方法处理综合评价指标集输入数据的基础上, 通过 Logistic 回归和神经网络(网络类型为多层感知器, 训练算法采用共轭梯度下降法)实现对基础研究成果多指标的综合评价并比较二法的评价效果。

关键词:神经网络; 共轭梯度下降法; Logistic 回归; 基础研究; 成果评价

中图分类号: G311

文献标识码: A

文章编号: 1001-7348(2005)09-0102-02

0 前言

科技成果评价是受多种因素、多维指标控制影响的复杂的社会活动, 是对多种性质不同的评价指标的综合评判。由于所有与评价结果有关的指标往往与评价结果之间并

的平均值时:

如上期与上年比较时价值创造的效率下降了, 意味着价值被破坏了。但声明一句, 不是任何价值创造的效率下降都是破坏性的。由投资引起的效率的短期下降是正常的。但是正如投资报酬率一样, 效率的下降不应该持续较长时间。

正如前面已说的, 价值破坏的第二种类型是效率比平均值低。每个公司由组织单位、附属公司、分支机构或相似的经济实体组成。单个结果加总就是总智力资本效率, 代表了公司的平均效率, 有低于平均效率的, 也有高于平均效率的。所有那些在平均值下运作的经济实体实际上是破坏价值, 理由是, 从公司层面来看, 同样的一个单位增

不完全是一种线性的函数关系, 因而可能不适于采用线性加权方法。Logistic 回归与人工神经网络都能够处理输出变量(应变量)为分类变量的非线性问题, 且对变量均不要求必须服从正态分布。Logistic 回归在已有数学模型基础上建立自变量与应变量之间的

加值要耗更多的资源。

由于过时的计量系统扭曲了经营业绩的评价, 今天的经济正处于危险境地。很多公司对其经济进步确信不疑, 而事实可能完全不同。通过运用新的指标, 公司能更清楚自己的经营能力, 从而对价值创造过程进行控制。

另外, 增加值和 IC 的效率再次将经济的微观和宏观层面联合起来, ICE 用同样的数据库来处理微观和宏观层面。换句话说, 公司与国家的 IC 平均效率相比, 将知道他们的效率怎样, 政府对其职责也将更清楚以便为经济的发展提供更宽松的环境。

最后, 智力资本作为一种资源的观念的转变和其价值创造效率也有社会学的成分,

函数关系。神经网络脱离传统的统计假设, 没有预先设定的数学模型, 靠网络自身来寻找输入变量与输出变量的映射关系, 具有很强的以任意精度逼近任意连续非线性函数的能力。本研究旨在比较二者在基础研究成果综合评价中的评价效果。

它强调了雇员作为价值创造者的重要性, 同时在投资者、管理者和雇员之间架起了沟通的桥梁, 也使组织和政府的联系更加紧密。

参考文献:

- [1] Marr.B.Gray.D.and Neely.A.Why do firms measure intellectual capital.Journal of intellectual capital, Vol.4, No.4.2003.
- [2] Drucker.P. California Management review, 1999.
- [3] Butterfield.H.(1949): The origins of modern science. London. 1-7.
- [4] 周炜炜, 华彬. 谁在创造财富, 谁在毁灭财富——2001年中国上市公司财富创造和毁灭排行榜[J]. 财经, 2002, (8): 36-39.

(责任编辑: 汪智勇)

收稿日期: 2005-01-11

基金项目: 辽宁省教育厅高等学校科学研究项目计划(20124346)

作者简介: 何钦成(1947-), 男, 中国医科大学副校长, 中国医科大学图书馆馆长, 信息管理与信息系统(医学)系主任, 教授, 主要研究方向为科研评估理论与技术; 黄亚明(1971-), 女, 中国医科大学信息管理与信息系统(医学)系讲师, 主要研究方向为科研评估理论与技术、信息计量学、信息资源分析与评价; 何苗(1975-), 女, 中国医科大学附属第一医院计算机室助理研究员, 主要研究方向为模式识别。

1 实验材料与与方法

本研究采用国家某部基础性研究成果奖(自然科学奖)全国推荐评奖项目 256 项作为样本。根据基础性研究的特征,拟定了如下 4 方面的评价指标:①反映研究工作起点的指标:承担课题得分。课题分为 9 类,即:“973”、“863”等国家重大课题、国家自然科学基金重点重大项目、国家杰出青年基金、国家自然科学基金面上项目、部委级课题、国际合作课题、省级课题、国内合作课题和厅市级课题。②反映研究工作量的指标:即项目执行时间(年)。③反映研究成果各种表现形式的指标:即在国内外期刊论文发表得分、参编专著得分(由专著主编中、专著副主编中、专著编委中署项目完成人的人数 3 种类别构成)、获专利得分(由发明专利、实用新型专利、外观设计专利项目数 3 种类别构成)及会议交流论文得分(由国际会议特邀报告及主题报告、国际会议论文、全国会议特邀报告、全国会议论文、部省级及以下会议论文篇次 5 种类别构成)④反映学术界的评价及成果学术影响力的指标,包括论文被检索工具收录得分(包括论文被 SCI、EI 和 ISTP 3 大检索工具收录、论文被其它国际性检索工具收录、论文被国内检索工具 CSDC 和 CSTPC 收录 3 种类别构成)、论文被国内外期刊引用得分、论文被国外专著引用得分、会议论文被引用得分、专著被引用得分等指标

各评价指标的量化按相关研究^[1-3]中的方法进行处理并进行 0-1 变换。输出变量或应变量为获奖的等级或由获奖等级所确定的赋值。本研究所采用的神经网络软件为 StatSoft 公司的 STATISTICA Neural Networks。为了获得较为理想的训练效果,网络类型选择多层感知器,训练算法采用共轭梯度下降法^[4]。将无量纲变换后的数据输入已建立的神经网络模型,输入变量共 11 个:课题得分、项目执行时间、会议交流论文得分、著书得分、专利得分、收录得分、被国外专著引用得分、专著被引用得分、会议论文被引用得分、论文发表得分和论文被引用得分。神经网络训练参数见表 1,输出变量处理见表 2。

按照两种评议思维方法进行模拟,第一种方法是一次性评价,即一次性评出各奖等(包括不获奖);第二种方法是两步评价法,即先评出获奖与否,再在获奖项目中评出获奖等级。样本按获奖等级和所属学科分两层,

表 1 神经网络参数表

	一次性评价 (全部项目获奖等级三值 输出)	二次性评价第一 步(全部项目按 获奖与否二值 输出)	二次性评价第二 步(获奖项目奖 励等级二值 输出)
样本量	256	113	256
样本随机分配			
Training Set(训练集)	178(70%)	79(70%)	178(70%)
Verification Set(校验集)	39(15%)	17(15%)	39(16%)
Test Set(测试集)	39(15%)	17(15%)	39(14%)
层数	3	3	3
隐含层单元数	47	8	18
Accept(Confidence Limits, 接纳可信限)	0	0.6641	0.6276
Reject(Confidence Limits, 拒绝可信限)	1	0.6641	0.6276
Activation Function(激活 函数)			
Layer 1(第一层)	Linear(线性)	Linear(线性)	Linear(线性)
Layer 2(第二层)	Logistic	Logistic	Logistic
Layer 3(第顶层)	Logistic	Logistic	Logistic

表 2 3 种方法按照两种评价方式对输出变量的赋值

	一次性评奖(三值输出)			两步评奖(二值输出)			
	一等奖	二等奖	未获奖	第一步		第二步	
				获奖	未获奖	一等奖	二等奖
神经网络	V1	V2	V3	V1	V2	V1	V2
Logistic 回归	2	1	0	1	0	1	0

按比例随机抽取样本,分配入训练集、校验集和测试集,三集相互独立。

训练数据,分别得到 3 个较为理想的网络。

应用 STATISTICA、SPSS 等统计分析软件进行 Logistic 回归。应用上述 11 个变量的原始得分(论文发表和论文被引用得分取平方根后的值)为自变量。

在神经网络和 logistic 回归中,输出变量按两种评价方式赋值(见表 2)。

最后对评价结果进行比较分析。

2 结果

2.1 神经网络运行结果

(1)在一次性评价(输出变量为三值)中,训练得到的较理想网络纳入全部 11 个输入变量。网络评价结果与实际同行评议结果的符合率为 71.9%。训练集、校验集和测试集的误差和正确分类比例分别为:0.3557, 0.7472;0.3704,0.7692;0.4274,0.6667。

(2)在两步评价中,评价各样本获奖与否时,训练所得网络纳入了全部 11 个输入变量。网络评价结果与实际同行评议结果的符合率为 76.7%。训练集、校验集和测试集

的误差和正确分类比例分别为:0.3761, 0.7667;0.3857,0.8095; 0.5086,0.7222。

(3)在两步评价中,评价获奖样本得一等奖或二等奖时,训练所得的网络纳入了全部 11 个输入变量。网络评价结果与实际同行评议结果的符合率为 76.1%。训练集、校验集和测试集的误差和正确分类比例分别为: 0.3718,0.7848; 0.3510,0.7059;0.5332, 0.7059。

2.2 Logistic 回归结果

(1)在应变量为三值输出的评价中,按 $P < 0.15$ 纳入论文被引用得分、项目执行时间、专著被引用得分、会议论

文得分和被国外专著引用得分 5 个变量。

建立 Logistic 回归预测概率模型:

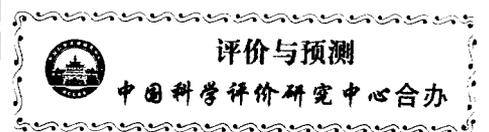
$$\hat{\pi}_{\text{不获奖}} = [\exp(4.130 - 0.107 \times \text{论文被引} - 0.121 \times \text{专著被引} - 0.092 \times \text{项目时间})] / [1 + \exp(4.130 - 0.107 \times \text{论文被引} - 0.121 \times \text{专著被引} - 0.092 \times \text{项目时间}) + \exp(1.743 - 0.029 \times \text{论文被引} - 0.409 \times \text{会议论文})]$$

$$\hat{\pi}_{\text{一等奖}} = [\exp(1.743 - 0.029 \times \text{论文被引} - 0.409 \times \text{会议论文})] / [1 + \exp(4.130 - 0.107 \times \text{论文被引} - 0.121 \times \text{专著被引} - 0.092 \times \text{项目时间}) + \exp(1.743 - 0.029 \times \text{论文被引} - 0.409 \times \text{会议论文})]$$

$$\hat{\pi}_{\text{二等奖}} = 1 / [1 + \exp(4.130 - 0.107 \times \text{论文被引} - 0.121 \times \text{专著被引} - 0.092 \times \text{项目时间}) + \exp(1.743 - 0.029 \times \text{论文被引} - 0.409 \times \text{会议论文})]$$

回归结果与同行评议结果的符合率为 65.3%。

(2)两步评价第一步即全部样本评价获奖与否(二值输出)时,进行 Logistic 前向逐步回归,按 $p < 0.15$ 水准纳入自变量 3 个:论文被引、课题执行时间和专著被引。



评价与预测

中国科学院评价研究中心主办

建立 Logistic 回归预测概率模型:

$$\hat{\pi}_i = \frac{\exp(-2.058 + 0.102 \times \text{论文被引得分} + 0.137 \times \text{专著被引得分} + 0.076 \times \text{项目执行时间})}{1 + \exp(-2.058 + 0.102 \times \text{论文被引得分} + 0.137 \times \text{专著被引得分} + 0.076 \times \text{项目执行时间})}$$

回归结果与同行评议结果的符合率为 72.2%。

(3) 两步评价第二步即获奖样本评价获奖等级(二值输出)时,进行 Logistic 前向逐步回归,按 $p < 0.15$ 水准纳入自变量 2 个:论文被引,会议发表论文。

建立 Logistic 回归预测概率模型:

$$\hat{\pi}_i = \frac{\exp(-1.260 + 0.038 \times \text{论文被引得分} + 0.371 \times \text{会议发表论文得分})}{1 + \exp(-1.260 + 0.038 \times \text{论文被引得分} + 0.371 \times \text{会议发表论文得分})}$$

回归结果与同行评议结果的符合率为 69.9%。

2.3 神经网络和 Logistic 回归结果比较

表 3 神经网络和 Logistic 回归结果与同行评价结果符合率(%)

	两步评奖(二值输出)	
	一次性评奖(三值输出)	第一步 第二步
神经网络	71.9	76.7 76.1
Logistic 回归	65.3	72.2 69.9

表 4 神经网络和 Logistic 回归结果——引入自变量(输入变量)

	两步评奖(二值输出)	
	一次性评奖(三值输出)	第一步 第二步
神经网络	11 个输入变量	11 个输入变量 11 个输入变量
Logistic 回归	论文被引,项目执行时间,专著被引,会议发表论文,被国外专著引	论文被引,课题执行时间,专著发表 论文被引,会议发表论文

3 讨论

周利锋等^[4]利用 25 例冠心病病人和 25 例正常人的资料进行分析,发现 BP 神经网络

方法在 Logistic 回归的变量独立性的假设不满足时,优于 Logistic 回归;又利用彼此独立的 3 个变量进行模拟,发现含隐含层的 BP 网络对样本的判别效果优于 Logistic 回归。本研究采用共轭梯度下降算法(Conjugate Gradient Decent)对基础研究成果申报项目进行了较大样本的神经网络训练(即网络模拟评估),进一步发现神经网络与 Logistic 回归法的评价即模拟效果之间存在差异。从评价符合率来看,神经网络的正确分类比率在各种分类评价中均高于 Logistic 法。从纳入的自变量来看,神经网络在各种评价中均纳入全部 11 个变量,Logistic 回归在一次性评价中纳入了论文被引得分、项目执行时间、专著被引得分、会议发表论文得分 4 个变量,这 4 个变量也是两种方法引入的共同变量,在两步评价的第一步纳入了论文被引得分、项目执行时间、专著被引得分,在第二步纳入了论文被引得分和会议交流论文得分。不论是应用何种方法和进行何种类别的评价,论文被引都是一个极其重要的被纳入影响因素,这正反映了对基础性研究成果水平评价中论文被同行引用是最重要的评价指标这一共识。项目执行时间和专著被引在次,这从另一个侧面反映出高水平的成果需要多年研究工作

的积累和应为同行所重视。Logistic 模拟纳入变量较少,可见该模型易受变量间共线性的影响,容易丢失对成果评价重要而有价值的信息,不能体现对成

果水平综合判断中对评价指标集利用的完整性和系统性。而神经网络方法不要求变量间相互独立,不受变量相关关系的影响,把评价指标集中的每个指标都作为输入值,并以此

给出对被评价成果水平的最终判断,能够保证成果评价中的系统性和完整性。

在本研究中,根据专家评议的思维方法,拟设了两种评价方式,一是一次性评出各个奖等的获奖项目,二是分两步评价,即先判定授奖与不授奖,然后在获奖组项目中评议出各授奖等级。实验结果前者符合率为 71.9%,后者两步符合率分别达到 76.7%和 76.1%。说明后一种网络评价方式与同行评议结果更为吻合。同时,Logistic 回归结果也呈现后者优于前者的趋势(两步评议符合率分别为 72.2%和 69.9%,一次性评议符合率 65.3%)。

参考文献:

- [1] 孟凯楠.可用于科研评价的一种定量分析方法——广义距离综合分析法[M].北京:红旗出版社,2000.176-179.
- [2] 何钦成,王孝宁,郭继军.期刊影响因子等级区域赋分方法的分析与改进[J].中国科技期刊研究,2003,(6):618-620.
- [3] 王孝宁,何苗,何钦成等.基于文献计量学研究方法的科技论文定量评价[J].科学学与科学技术管理,2004,(4):15-18.
- [4] 何钦成,王孝宁,韩大勇,郭继军.TOPSIS 法在科技人员业绩评价中的应用[J].科研管理,2004,(4):16-22.
- [5] 黄亚明,何苗,何钦成.神经网络在基础研究成果综合评价中的应用[A].中国医科大学硕士学位论文,2004.
- [6] 谢政,李建平,汤泽莹.非线性最优化[M].北京:国防科技大学出版社,2003.183-184.
- [7] 张铁,闫家斌.数值分析[M].北京:冶金工业出版社,1986.52-56.
- [8] 周利锋,高尔生,金玉焕.BP 神经网络与 Logistic 回归对经初探[J].中国卫生统计,1998,(1):1-4.

(责任编辑:汪智勇)

Comparative Study of Artificial Neural Network and Logistic Regression in Evaluating Basic Research Achievement

Abstract: To construct a comprehensive evaluating system of basic scientific and technologic achievements in which multiple indexes should be comprised. The input and output data were dealt with methods of scientometrics, mathematics and statistics. And the systemic differentiation was implemented with Logistic Regression and artificial neural network(Multilayer Perceptrons as network type and Conjugate Gradient Decent as training algorithm). The results of evaluation with the two methods were compared and analyzed.

Key words: neural network; conjugate gradient decent; logistic regression; basic research; achievement evaluation