

基于神经网络的高校科研项目前景预测系统研究与设计

王汉元, 姜 英

(大连东软信息学院软件工程系, 辽宁 大连 116023)
✉wanghanyuan16@dnui.edu.cn; jiangying@neusoft.edu.cn



摘 要: 为了对高校科研项目前景进行有效预测, 进一步提升科研成果转化率, 本文提出了基于神经网络结合科研项目共享的网络平台, 构建可靠的科研项目前景预测系统。通过将各类大量科研项目的特征参数化, 利用BP(反向传播)神经网络的非线性映射能力、自适应能力和对离散数据的泛化能力生成模型后, 用于从多维度对一个新项目生成可靠的前景预测。实验表明, BP神经网络算法在大量学习经过预处理的样本后所产生的预测结果是具有较高准确性的, 且能够随着新的样本输入不断更新和适应, 因此该方法具有较强的可行性。

关键词: 高校科研项目; 前景预测; 神经网络

中图分类号: TP183 **文献标识码:** A

Research and Design of Prospect Forecast System of University Scientific Research Projects based on Neural Network

WANG Hanyuan, JIANG Ying

(Software Engineering Department, Dalian Neusoft Institute of Information, Dalian 116023, China)
✉wanghanyuan16@dnui.edu.cn; jiangying@neusoft.edu.cn

Abstract: This paper proposes to build a reliable prospect forecast system for scientific research projects in order to effectively predict prospects of scientific research projects in universities and further improve conversion rate of scientific research results. The system is based on neural networks and a network platform shared by scientific research projects. A model is generated by parameterizing characteristics of a large number of various scientific research projects, using the non-linear mapping ability, adaptive ability and generalization ability of BP (Back Propagation) neural network. Then, the model is used for reliable forecast of a new project from multiple dimensions. Experiments show that the prediction results, produced by the BP neural network algorithm after learning a large number of pre-processed samples, have high accuracy, and can be updated and adapted with new sample input. Therefore, this method is highly feasible.

Keywords: university scientific research projects; prospect prediction; neural network

1 引言(Introduction)

在我国, 科研成果转化率低已是老生常谈的话题, 2018年4月18日国务院常务会议确定了“加大对科技成果转化的政策激励”的政策方针并提出了若干办法。但由于科研评价机制不合理、选题立项重复度高、科研资金管理存在漏洞^[1]等问题, 导致科研成果转化率低的现象至今依然存在。提高科研成果转化率, 企业的参与是至关重要的, 然而企业为了规避风险很注重项目的成熟度, 偏向于投资前景已清晰的项目。

这样便导致很多优秀的项目由于找不到投资者止步在“原始创新成果”阶段, 失去了产业化、商品化的机会。而各个高校的科研项目展示页面更注重展示本校科研项目的创新性和技术高度, 而忽略了展示方案的实现成本及产业化的难度^[2], 这让企业难以理解其价值。

本文中所研究的采用SSM+MongoDB技术搭建的基于BP神经网络(反向传播神经网络)算法的高校科研项目前景预测系统, 可实现对科研项目进行可靠的前景预测, 帮助正在进行

中的科研项目更直观地展示自身价值,降低企业投资风险,提高科研成果的转化率^[3]。

2 技术简介(Technical brief)

基于BP神经网络算法的高校科研项目前景预测系统的技术选型使用SSM框架,数据库采用MySQL+MongoDB的形式,核心功能的实现使用了BP神经网络算法。

2.1 BP神经网络

BP神经网络是19世纪80年代由David Rumelhart等科学家提出的概念,是一种对感知机网络的改进算法,解决了简单感知器不能解决的异或(Exclusive OR, XOR)和非线性学习等问题。BP神经网络在训练过程中通过反复调整系统中各种参数的值形成自己的学习能力,并将这种能力应用于更多样本的分类,因此被广泛应用于分类和识别领域。从收集到的资料可以得知:神经网络算法是一种对绩效评价比较理想的方法,而且很多开发者在不同领域中应用这种方法,如企业绩效评价、动态供应链绩效评价、高校院系绩效评价等。其学习流程可简单概括为:

- (1)输入经过预处理的样本数据。
- (2)乘以权重,增加偏置并激活,逐层传递。
- (3)得到预测值并对比真实值得到损失值。
- (4)利用损失值对偏置和权重求偏导。
- (5)利用梯度下降的方法更新参数。
- (6)循环以上流程直到损失值达到标准。
- (7)训练完成。

BP神经网络的优势在于:(1)非线性映射能力:BP神经网络实质上实现了一个从输入到输出的映射功能,数学理论证明三层的神经网络就能够以任意精度逼近任何非线性连续函数。这使得其特别适于求解内部机制复杂的问题。(2)容错能力:BP神经网络在其局部的或者部分神经元受到破坏后,对全局的训练结果不会造成很大的影响,也就是说即使系统在受到局部损伤时还是可以正常工作的,即BP神经网络具有一定的容错能力。但是BP神经网络仍有一些缺陷,比如:容易陷入局部极小值;神经网络收敛速度缓慢;神经网络结构选择没有统一的解决方案,只能凭经验判断;神经网络预测能力和训练能力的矛盾,可能导致“过拟合”的情况。这些缺陷有的已经有了解决方案,而有些缺陷的解决方案还有待探索。

2.2 SSM框架

SSM框架是一种常用且经典的以MVC模式搭建网站的框架组合,其中包含三个框架:Spring、SpringMVC、MyBatis。Spring负责管理集成的框架、呈现切面、管理事务处理和Bean的生命周期;SpringMVC负责转发请求、传递页面中动态数据和业务处理;MyBatis负责和MySQL数据库的交互,将表转换成类,将字段转换成类的属性以方便程序员对数据库的CRUD^[4]。

2.3 MongoDB数据库

MongoDB是一个非关系型数据库,在数据的存储结构上却与关系型数据库和非关系型数据库都有相似之处。即使之

前对NoSQL的概念知之甚少的开发者也能快速上手。其优势在于清晰、便于理解的表结构,松散的存储方式可以方便地存储一些关系复杂的数据,而其作为非关系型数据库也具有大数据量处理、高性能读写的先天特性^[5]。

3 BP神经网络在核心功能模块中的应用 (Application of BP neural network in core function module)

使用BP神经网络进行前景预测的原因在于,通过研究发现,在特定类别的科研项目中,最终成功转化的科研项目在某些特征上具有共性,而不甚成功的科研项目同样在某些特征上具有另一些共性^[6]。这种数据关系与BP神经网络的工作模式非常符合,因此选用BP神经网络来完成前景预测功能。

由于神经网络的输入数据需要经过预处理,而计算机难以实现对科研项目的特征进行评估,因此在设计时增加了评论功能,通过这个功能的设计帮助富有经验的用户以规范的格式对项目进行全面且理性的评价。在经过一段时间,取得评论达到一定数量后,对评论取均值即得到一个项目信息较为准确客观的特征数据(取得了样本的特征值);而在项目结束后,会有合作的企业或者专业的测评人员按照评分标准给项目进行最终评分(取得了样本的真实值)。这样便收集到了训练神经网络所需样本的所有数据,经过预处理后的部分样本数据如表1所示。

表1 训练样本数据表

Tab.1 Data table of training samples

样本编号	特征1	特征2	特征3	特征4	特征5	真实值
1	0.75	1.0	1.25	1.0	0.75	A
2	1.75	2.0	1.75	2.0	1.75	A
3	1.75	1.75	1.25	1.75	1.25	A
4	2.0	2.0	1.25	0.5	1.25	A
5	1.25	2.0	1.0	1.0	1.25	A

训练神经网络时输入数据必须格式统一,因此经过研究设定了描述一个科研项目的五项特征:“选题”“经费”“人员”“可行性”“市场需求”。从这些角度评价可以全面地概括一个科研项目的各项特征,这五项特征的值对应BP神经网络输入层的五个神经元。

由于采用上述特征值收集方式使得特征值相差并不大,因此BP神经网络的传播方式采用Log-Sigmoid激活函数,如公式(1)所示:

$$y = 1 / (1 + \exp(-x)) \quad (1)$$

Math.exp为Java内用于返回自然数底数e的参数次方。而为了避免出现由于神经网络层数过深,在反向传播时出现梯度消失的情况,隐含层(Hidden-layer)采用三层神经元。隐含层的神经元个数由输入层和输出层神经元个数决定,具体公式如公式(2)所示:

$$hdNum = \sqrt{inNum + outNum} + ADJUST \quad (2)$$

Math.sqrt()为Java内获取参数正平方根的方法,hdNum

为隐含层神经元个数, inNum为输入层神经元个数, outNum为输出层神经元个数, ADJUST为节点调节常数, 这样的设计保证了在特征数量增加时程序的可扩展性。

在确定了神经网络的形状之后是规定隐含层神经元的计算规则, 如公式(3)所示:

$$a_{[l]} = \sigma(w_{[l]}a_{[l-1]} + b_{[l]}) \quad (3)$$

其中, $a_{[l]}$ 为 l 层神经元的输出值, σ 为激活函数, $w_{[l]}$ 为 l 层神经元的权重, $a_{[l-1]}$ 为 $l-1$ 层神经元的输出值, $b_{[l]}$ 为 l 层的偏置值。在得到最终输出值后与真实值比较取得损失值的损失函数, 如公式(4)所示:

$$L(a, y) = -\frac{1}{m} \times \left(\sum_{i=0}^m (y_{(i)} \times \log(a_{(i)}) + (1 - y_{(i)}) \times \log(1 - a_{(i)})) \right) \quad (4)$$

其中, a 为预测值, y 为真实值, m 为样本个数。随后使用梯度下降的方式更新网络中各节点的权值与偏置, 并再次训练, 重复直至损失值达到预设的标准^[7]。

4 系统设计(System design)

整个网站平台在SSM的框架下搭建, 在此介绍用于实现科研项目前景预测的主要功能模块的设计。

4.1 功能模块结构

系统所涉及的主要功能模块包含科研项目信息模块、评论模块和评分模块。科研项目模块主要包含的功能为创建和修改科研项目信息、查看项目详情和项目前景预测。评论模块和评分模块包含的功能为记录用户对科研项目各项特征的描述, 这些数据将为科研项目信息模块使用BP神经网络进行前景预测提供重要的数据支撑。

4.2 核心模块的工作流程

首先, 高校科研人员用户会使用平台上传科研项目信息, 科研项目信息模块会对这些信息进行整理, 之后存入数据库, 然后其他用户在浏览科研项目信息后留下评论, 这些评论由若干标签构成。这些标签被预先设置好放在编写评论界面的下方备选, 标签的内容为科研项目在若干特征上的表现, 如描述项目“经费”特征的标签有“经费不足”“经费合适”“过分充裕”; 描述项目“选题”特征的标签有“意义重大”“价值一般”“毫无意义”。此外还有描述其他多个特征的若干标签。标签界面设计如图1所示。

可行性	市场需求	经费	选题	人员
☐ 可行性高	☐ 契合需求	☐ 过分充裕	☐ 意义重大	☐ 人员优秀
☒ 一般	☒ 竞争激烈	☒ 经费合适	☒ 价值一般	☒ 人员合适
☐ 难度较大	☐ 没有市场	☐ 经费不足	☐ 毫无意义	☐ 不可靠

图1 标签界面设计

Fig.1 Design of label interface

这些标签被赋予不同的分值以衡量项目在若干特征上的表现。采用这种标签的评论方式可以借助评论者的知识对项目给出相对客观、全面且标准化的描述。项目与企业进行合作后, 企业用户会根据评分标准给出一个科研项目的最终评价。一个项目在收集若干评论数据后, 这些评论将取均值作

为该项目的特征值, 而最终评价则作为项目的真实值。在获得足够的样本后, 这些数据被用于训练BP神经网络和测试其准确度。BP神经网络训练完成后就可以通过一个科研项目的评论数据预测其前景。

4.3 系统测试结果

在测试中使用的训练集包含160条样本, 测试集包含40条样本, 训练完成后的BP神经网络预测准确率最高可达到95.0%, 该次训练结果如图2所示。

```
Train End!
count6
训练集的数量: 160
测试集的数量: 40
分类正确的数量: 38
算法的分类正确率为: 0.95
分类结果储存在: c:\Users\Administrator\Desktop\frameANow\WebContent\dataCollection\tmpResult.txt
```

图2 BP神经网络训练结果

Fig.2 The training result of BP neural network

4.4 数据存储

由于MySQL作为关系型数据库能够保证数据的ACID特性, 因此用于存储用户的账户相关信息; 而MongoDB则用于存储科研项目的相关信息, 如科研项目的介绍、附件、其他用户的评论等需要频繁进行CURD操作的数据。

5 结论(Conclusion)

由于系统所需数据需要平台正式上线并在具有相当影响力后才能取得, 真实数据难以收集, 且本文中对科研项目前景预测的设计仅为给相关工作者提供参考思路, 因此在测试设计有效性时使用了模拟数据。这样的结果说明BP神经网络在理想的条件下并不能百分之百的实现实验样本的准确分类, 但是实验结果中该算法展现较高的准确率已然表明, 在高校科研信息共享平台的开发中, BP神经网络应用于前景预测是具有较高可行性的。

参考文献(References)

- [1] 赵红梅.基于新形势下的高新技术企业科研经费管理探究[J]. 财经界(学术版),2018(4):84-86.
- [2] 于晓棠.简述高校科研成果转化存在的问题及对策[J].科技资讯,2020,18(12):217-218.
- [3] 高敏.基于BP神经网络的绩效评价应用研究[D].兰州:西北师范大学,2016.
- [4] 王燕贞,沈毅波.基于SSM框架的高校学生综合测评系统设计与实现[J].通化师范学院学报,2020,41(04):58-63.
- [5] 黄承明.基于MongoDB文档模型的教学资源数据的建模研究[J].软件工程,2020,23(05):46-49.
- [6] 李言荣.科研评价要减少“人”的因素[N].中国科学报,2020-06-03(001).
- [7] D. J. C. MacKay. A practical Bayesian framework for back-propagation networks[J]. Advances in Neural Information Processing Systems, 1992, 4(3):448-472.

作者简介:

王汉元(1998-),男,本科生.研究领域:软件工程,Java.

姜英(1978-),女,硕士,副教授.研究领域:软件工程.