**学号：2015200801**

**北京化工大学**

**硕士研究生学位论文开题报告**

**论文题目：ELM神经网络结构自组织的研究与应用**

**学 院 名 称：信息科学与技术学院**

**专 业：计算机科学与技术**

**研究生姓名： 李雅楠**

**导师姓名：耿志强**

**开题日期：2016年9月23日**

|  |  |  |
| --- | --- | --- |
| **考核**  **成绩** |  | |
| **审核**  **小组**  **成员**  **以及**  **职称** | **姓 名** | **职 称** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

目录

[1. 课题来源及项目名称 3](#_Toc462270863)

[2. 文献综述 3](#_Toc462270864)

[2.1 研究背景 3](#_Toc462270865)

[2.2. ELM神经网络理论研究情况 4](#_Toc462270866)

[2.3. 中心极限定理 6](#_Toc462270867)

[2.4. 信息熵理论 6](#_Toc462270868)

[2.5. 余弦相似定理 7](#_Toc462270869)

[2.6. 参考文献 8](#_Toc462270870)

[3. 研究计划 10](#_Toc462270871)

[3.1. 选题的目的及意义 10](#_Toc462270872)

[3.2. 主要研究内容 11](#_Toc462270873)

[3.3. 研究方案 11](#_Toc462270874)

[3.3.1. 技术方案 11](#_Toc462270875)

[3.3.2. 实施方案所需要的条件 11](#_Toc462270876)

[3.3.3. 拟解决的关键问题 12](#_Toc462270877)

[3.4. 课题难点分析 12](#_Toc462270878)

[3.5. 预期研究成果及创新点 12](#_Toc462270879)

[3.6. 工作计划进度 12](#_Toc462270880)

# 课题来源及项目名称

课题来源：国家自然科学基金项目《基于DEA的复杂化工过程能效分析与预测方法研究》61603025

项目名称：ELM神经网络结构自组织的研究与应用

# 文献综述

## 研究背景

神经网络是一门新兴的交叉学科，始于20世纪40年代，是人类智能研究的重要组成部分。人工神经网络是对人类大脑系统的一阶特性的描述。简单地讲，它是一个数学模型，可以用电子线路来实现，也可以用计算机程序来模拟，是人工智能研究的一种方法。它已广泛应用于脑科学，神经科学，认知科学，心理学，计算机科学，数学和物理学等领域。神经网络模型的最重要的优点是其灵活性，以模拟化学特性的非线性特性和能力通过学习输入和输出数据之间的关系来近似与高维数据的任何功能和技术的一个可接受的水平预测他们精度。

在1986年，Rumelhart和McCelland为首的科学家提出了BP（Back Propagation）神经网络。BP神经网络具有较强的非线性映射，识别，分类和非线性处理能力。Kou等人提出基于BP神经网络特征选择，利用AIS（fsbp-ais）方法来确定BP神经网络的连接权值的模型，可应用于预测采摘人员的位置，也可以应用于生产领域，预测农作物的产量。Deosarkar等人提出基于BP神经网络的网络模型成功地用于磁铁矿矿石泥浆的有效粘度预测。但是BP神经网络自身也有一些缺陷和不足。首先，BP神经网络收敛速度慢，对于一些复杂的问题，需要的训练时间非常长；其次BP神经网络采用梯度下降法，容易产生局部最优解；再次网络隐含层数和单元数的选择尚无理论指导，因此需要根据经验或者反复实验确定；最后，BP网络的学习和记忆具有不稳定性。1985年，鲍威尔提出径向基函数（RBF）多变量插值方法，1988年，Broomhead等人设计并提出了RBF神经网络。RBF是一种高性能的前馈神经网络，可以解决BP神经网络的局部最优解问题。已经成功应用于非线性函数逼近，时间序列分析，数据分类，模式识别，信息处理，图像处理，系统建模，控制和故障诊断等问题。但是RBF神经网络容易丢失数据，无法解释基础的推理过程。因此学习算法需要进一步改进和提高。

南洋理工大学黄广斌在2004年提出的Extreme Learning Machine（ELM，极限学习机），是一种单隐层前馈神经网络（SLFN）学习算法。相对于传统前馈神经网络训练速度慢，容易陷入局部极小值点，学习率的选择敏感等缺点，ELM 算法随机产生输入层与隐含层的连接权值及隐含层神经元的阈值，且在训练过程中无需调整，只需要设置隐含层神经元的个数，便可以获得唯一的最优解。与之前的传统训练方法相比，ELM 方法具有学习速度快，泛化性能好等优点。但是ELM初始权值随机设置往往不能获得最优的网络结构，如有些节点对网络性能的作用影响是不可以忽略不计的。而且由于样本数量的不同，需要人工多次实验确定隐含层的节点数，无法根据样本自组织确定，且在实验过程中，容易出现过拟合的问题。所以以后的研究方向主要就ELM这些缺陷进行改进。

在研究过程中，结合信息熵，中心极限定理和余弦相似定理的方法，对ELM网络进行改进。中心极限定理中的协方差在科学研究中特别是农业和金融的领域有着非常重要的应用。Gao等人提出将协方差分析应用在油菜田实验中；Turtle等人提出了投资组合形式的协方差估计方法，协方差可以消除不同测试结果间的影响，提高了实验结果的可靠性。熵是确定基于概率论确定信息不确定性的一个重要方法，广泛应用于采样和通信领域。Dehmer论证了图片的信息熵在生物学，化学和社会学中起着重要的作用。Zou等人介绍了一种新的权重评价方法，通过对模糊综合评价方法困难程度的考虑，在计算过程中，考虑了多个因素和评价对象之间的关系。Chen等人将熵应用于人工神经网络中，并对降雨-径流模拟的结果进行了预测。在数学中，运用余弦相似定理通过测量两个分析相似性，在各个领域中有着非常广泛的应用。相似性度量是确定两个对象之间相似程度的一个重要工具。Lin等人提出利用余弦相似定理，来获得在文档中某个关键字出现的频率。

## ELM神经网络理论研究情况

传统的神经网络结构如下图一所示，由输入层，隐含层和输出层组成。输入层与隐含层，隐含层与输出层神经元间全连接。其中，输入层有n个神经元，对应n个输入变量，隐含层有l个神经元，输出层有m 个神经元，对应m个输出变量。在不失一般性的情况下，设输入层与隐含层间的连接权值W 为：。

其中，表示输入层第i个神经元与隐含层第j个神经元的连接权值。

设隐含层与输出层间的连接权值为：

其中，表示隐含层第j个神经元与输出层第k个神经元间的连接权值。

设具有Q个样本的训练集输入矩阵x和输出矩阵Y分别为：

 

E:\陈杰论文\图\我自己的\tif\ELM.tif

极限学习机是黄广斌提出的一种单隐层神经网络，具有快速学习的特点。对于单隐层神经网络，可以初始化输入的连接权值和偏差，并得到相应的输出连接权值。产生最佳的解决方案，更快的学习速度和较好的泛化性能。ELM广泛应用于能效分析和神经计算。假设有N个随即样本, ，，。对于一个有L个隐层节点的单隐层神经网络可以表示为：

其中，为激活函数，为输入权重，为输出权重，是第i个隐层单元的偏置。表示和的内积。单隐层神经网络学习的目标是使得输出的误差最小，可以表示为:。即存在，和，使得。可以表示矩阵为：。其中，是隐含层节点的输出，为输出权重，为期望输出。

，，。

为了能够训练单隐层神经网络，我们希望得到，和，使得：



其中，，这等价于最小化损失函数：。传统的一些基于梯度下降的算法，可以用来求解这样的问题，但是基本的基于梯度的学习算法需要在迭代过程中调整所有的参数。而在ELM算法中，一旦输入权重和隐层的偏置被随机确定，隐层的输出矩阵就被唯一确定。训练单隐层神经网络可以转化为求解一个线性系统。并且输出权重可以被确定为。其中，是矩阵的Moore-Penrose广义逆。且可证明求得的解的范数是最小的并且唯一。

## 中心极限定理

大数定律揭示了大量随机变量的平均结果，但没有涉及到随机变量的分布的问题。而中心极限定理说明的是在一定条件下，大量独立随机变量的平均数是以正态分布为极限的。中心极限定理是概率论中最著名的结果之一。它提出，大量的独立随机变量之和具有近似于正态的分布。因此，它不仅提供了计算独立随机变量之和的近似概率的简单方法，而且有助于解释为什么有很多自然群体的经验频率呈现出钟形(即正态)曲线这一事实[9]，因此中心极限定理这个结论使正态分布在数理统计中具有很重要的地位，也使正态分布有了广泛的应用。

在这篇文章中，将概率论应用到改进神经网络中，高斯分布是分析系统状态的重要工具[10]。SLFN的结构由高斯分布来根据样品决定的。高斯分布有着极其广泛的实际背景，在生产与科学实验中，有很多的随机变量的概率分布都可以近似地用高斯分布来描述[11]。中心极限定理是概率论中讨论随机变量序列部分和分布渐进于正态分布的一类定理。一般来说，如果一个量是由许多微小的独立随机因素影响的结果，那么就可以认为这个量具有正态分布。

## 信息熵理论

1948年，香农提出了“信息熵”的概念，解决了对信息的量化度量问题[12~14]。信息熵是一个数学上颇为抽象的概念，在这里不妨把信息熵理解成某种特定信息的出现概率。对于任意一个随机变量，它的熵的定义如下：

变量的不确定性越大，熵也就越大，把它搞清楚所需要的信息量也就越大。

信源的平均不定度。在信息论中信源输出是随机量[15]，因而其不定度可以用概率分布来度量[16]。记，这里为信源取第i个符号的概率。称为信源的信息熵。

可以从数学上加以证明，只要满足下列三个条件：

1. 连续性：是的连续函数；
2. 对称性：与的排列次序无关；
3. 可加性：若，且，则有；则一定有下列唯一表达形式：

其中为正整数，一般取，它是信息熵的最基本表达式。

信息熵的单位与公式中对数的底有关。最常用的是以2为底，单位为比特(bit)；在理论推导中常采用以e为底，单位为奈特(Nat)；还可以采用其他的底和单位，并可进行互换。

信息熵除了上述三条基本性质外，还具有一系列重要性质，其中最主要的有：

1. 非负性：；
2. 确定性：；
3. 扩张性：；
4. 极值性：；这里；
5. 上凸性：，式中。

## 余弦相似定理

相似度度量（Similarity），即计算个体间的相似程度，相似度度量的值越小，说明个体间相似度越小，相似度的值越大说明个体差异越大[17~20]。余弦相似度用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小[21]。余弦值越接近1，就表明夹角越接近0度，也就是两个向量越相似，这就叫"余弦相似性"[22]。

两个向量a,b的夹角很大可以说a向量和b向量有很低的的相似性，或者说a和b向量代表的文本基本不相似。那么是否可以用两个向量的夹角大小的函数值来计算个体的相似度呢？向量空间余弦相似度理论就是基于上述来计算个体相似度的一种方法[23~25]。下面做详细的推理过程分析。

三角形中边a和b的夹角 的余弦计算公式为：

在向量表示的三角形中，假设a向量是（x1, y1），b向量是(x2, y2)，那么可以将余弦定理改写成下面的形式：向量a和向量b的夹角 的余弦计算如下：



如果向量a和b不是二维而是n维，上述余弦的计算法仍然正确。假定a和b是两个n维向量，则a与b的夹角的余弦等于：



## 2.6. 参考文献

[1] Huang G B, Zhou H, Ding X, et al. “Extreme learning machine for regression and multiclass classification,” IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, vol.42(42),pp.513-29,2012.

[2] Huang G B, Zhu Q Y, Siew C K, “Extreme learning machine: a new learning scheme of feedforward neural networks,” Proc.int.joint Conf.neural Netw, vol.2,pp.985-990,2004.

[3] Huang G B, Zhu Q Y, Siew C K, “Extreme learning machine: Theory and applications,” Neurocomputing vol.70(1–3),pp.489-501,2006.

[4] Aghbashlo M, Shamshirband S, Tabatabaei M, et al. “The use of ELM-WT (extreme learning machine with wavelet transform algorithm) to predict exergetic performance of a DI diesel engine running on diesel/biodiesel blends containing polymer waste,” Energy94,pp.443-456, 2016.

[5] Li S, Goel L, Wang P, “An ensemble approach for short-term load forecasting by extreme learning machine,” Applied Energy, vol.170,pp.22-29,2016.

[6] Kariminia S, Shamshirband S, Motamedi S, et al. “A systematic extreme learning machine approach to analyze visitors׳ thermal comfort at a public urban space,” Renewable & Sustainable Energy Reviews,vol.58,pp.751-760, 2016.

[7] Wang S, Deng C, Lin W, et al. “A novel NMF-based image quality assessment metric using extreme learning machine,” IEEE transactions on cybernetics, pp.255-258,2016.

[8] Liu Q, Yin J, Leung V C M, et al. “Applying a new localized generalization error model to design neural networks trained with extreme learning machine,” Neural Computing & Applications, vol.27(1),pp.59-66,2016.

[9] Gao T J, Qiang S, Song X L, et al. “Utilizing statistical control in herbicide efficacy evaluation,” Agrochemicals, 2006.

[10] Turtle H J, Wang K. “The benefits of improved covariance estimation,” Journal of Empirical Finance, vol.37, pp.233-246, 2016.

[11] Hilty L M, “Information, Entropy and Environmental Problems,” Current Rheumatology Reports, vol.16(5),pp.1-9,2014.

[12] Dehmer M, Mowshowitz A, “A history of graph entropy measures,” Information Sciences, vol.181(1),pp.57-78,2011.

[13] Zou Z H, Yun Y, Sun J N, “Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment,” Journal of Environmental Sciences, vol.18(5),pp.1020-1023,2006.

[14] Liu L, Zhou J, An X, et al. “Using fuzzy theory and information entropy for water quality assessment in Three Gorges region, China,” Expert Systems with Applications, vol.37(3),pp.2517-2521,2010.

[15] Chen L, Singh V P, Guo S, et al. “ Copula entropy coupled with artificial neural network for rainfall–runoff simulation,” Stochastic Environmental Research & Risk Assessment, vol.28(7),pp.1755-1767,2014.

[16] Lam K C, Lam C Y Y K Y, “An Artificial Neural Network and Entropy Model for Residential Property Price Forecasting in Hong Kong,” Journal of Property Research, vol.25(4),pp.321-342,2008.

[17] Ye J, “Cosine similarity measures for intuitionistic fuzzy sets and their applications,” Mathematical & Computer Modelling, 53(1-2):91-97,2011.

[18] Fawaz S. Al-Anzi, Dia AbuZeina, “Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing,” Journal of King Saud University - Computer and Information Sciences.

[19] Xia P, Zhang L, Li F, “Learning similarity with cosine similarity ensemble,” Information Sciences, vol.307,pp.39-52,2015.

[20] Hou X N, Ding S H, Ma L Z, et al. “Similarity metric learning for face verification using sigmoid decision function, ”Visual Computer, pp.1-12,2015.

[21] Lin Z C, Wu D W, Hong G E, “Combination of improved cosine similarity and patent attribution probability method to judge the attribution of related patents of hydrolysis substrate fabrication process, ”Advanced Engineering Informatics, vol.30(1),pp.26-38,2016.

[22] Fawaz S. Al-Anzi, Dia AbuZeina, “Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing,” Journal of King Saud University - Computer and Information Sciences.

[23] Nguyen H V, Bai L, “Cosine Similarity Metric Learning for Face Verification”// Computer Vision - ACCV 2010 -, Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers,pp.709-720,2010.

[24] Han Y, Geng Z, Zhu Q, et al, “Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry,” Energy, vol.83,pp.685-695,2015.

[25]Han Y M, Geng Z Q, Liu Q Y, “ Energy Efficiency Evaluation Based on Data Envelopment Analysis Integrated Analytic Hierarchy Process in Ethylene Production,” Chinese Journal of Chemical Engineering, vol.22(11-12),pp.1279-1284,2014.

# 3. 研究计划

## 3.1. 选题的目的及意义

在人工神经网络研究方面，国内的起步较晚。近年来，基于人工神经网络的各种应用研究十分热门，涌现了大量专著和文献研究报告。期刊文献中有相当多的关于结合人工神经网络进行各领域应用的研究，也有很多文献是关于利用遗传算法、专家网络、进化算法、模糊算法等对人工神经网络进行改造，产生一些新的神经网络学习训练算法。神经网络日益成为当代科技领域中方兴未艾的热点。

ELM的参数设置简单，如最初提出的极限学习机算法只需要设置好合适的隐层节点就能取得良好的性能。并且其计算复杂度很低，运算速度很快，无需像BP网络那样迭代求解，也无需像支持向量机那样求解复杂的二次优化。ELM不会像BP神经网络那样容易陷入局部最优解，这是因为它求解过程是一个凸优化的过程，因而具有良好的泛化性能，并且采用了神经网络的框架因而可以很容易处理多分类问题。虽然ELM优点较多，但是初始权值随机设置往往不能获得最优的网络结构，如有些节点对网络性能的作用影响是不可以忽略不计的。而且由于样本数量的不同，需要人工多次实验确定隐含层的节点数，无法根据样本自组织确定，且在实验过程中，容易出现过拟合的问题。

本课题旨在结合中心极限定理，自组织确定神经网络的隐含层节点数。并且结合熵和余弦定理，设置网络初始权值和隐含层到输出层的权值。最终，为ELM神经网络结构和参数提供一种高效、简洁的理论支持。

## 3.2. 主要研究内容

1）研究ELM神经网络的结构特征，算法改进，算法结构优化等理论知识，改进现有的ELM神经网络的结构设计方法，借助中心极限定理的研究成果，对ELM的网络学习性能进行优化。

2）借助信息熵的知识和余弦定理，研究ELM初始连接权值的设定，实现优化ELM的网络结构，提高分类和预测的精度。使其结构的设计更加合理。

3）在UCI数据集上训练并测试ELM网络，验证改进的ELM结构能否获得良好的训练效果，并且应用在石化生产的产量预测上。

## 3.3. 研究方案

### 3.3.1. 技术方案

1）针对第一点研究内容，先研究传统ELM的结构特征，然后深入学习其他学者提出的的对ELM进行算法改进和结构优化的论文。研究概率论的知识，对ELM的结构进行改进。

2）深入透彻学习信息熵，概率论，余弦相似定理应用于相似性检测的理论知识。研究现有的信息熵和神经网络尤其是和ELM结合的文献资料，着手结合信息熵对ELM初始权值的改进。然后学习余弦相似定理相关的文献，应用在隐含层到输出层连接权值的确定上。并应用UCI数据集测试其分类精度和预测的准确度。

3）将研究结果应用在石化生产上，使用乙烯数据进行测试。使其具有更好的建模精度，拓展神经网络技术在化工领域的应用，知道乙烯的生产预测。

### 3.3.2. 实施方案所需要的条件

1) 所需开源环境如Python开源库Theano，deepnet等，也可使用matlab库

2) 测试数据集如UCI数据集、乙烯数据等

### 3.3.3. 拟解决的关键问题

1）自组织设计模型结构，解决ELM神经网络拓扑结构隐含层节点数无法根据样本实现自组织网络结构，以至于对学习性能产生不利影响问题。

2）结合熵和余弦定理，得到神经网络输入层到隐含层，隐含层到输出层的连接权值。提高预测的准确性如何更好的应用在石化生产的产量预测上。

## 3.4. 课题难点分析

1）ELM神经网络的结构，隐含层节点数的设定都将影响网络在不同实际问题中的学习性能，如何找到有效的结构构建准则和能够表达样本数据的隐含层节点的构造方法也是本课题的难点之一。

2）ELM神经网络各层间的连接权值也会对最后的预测结果产生影响，如何利用熵和余弦定理确定权值，提高预测准确性。以及学习过程中的样本数量，样本的选取对学习性能也有很大的影响，找到有效的调整规则也是本课题的难点。

## 3.5. 预期研究成果及创新点

预期研究成果：

1）提出一种结合中心极限定理改进的ELM神经网络，能够实现网络结构的自组织，根据样本确定网络隐含层节点数，从而对不同信息数据的处理需求有良好的学习效果。

2）结合熵和余弦相似定理，确定输入层到隐含层，隐含层到输出层之间的连接权值，使得预测的结构更为准确。

3）验证所提出优化算法在石化生产领域的可行性。

4）在研究的工作基础上，撰写硕士论文，发表学术文章1~2篇。

创新点：

1）结合中心极限定理，根据样本，自组织确定ELM的隐含层节点数。

2）结合熵和余弦相似定理，确定输入层到隐含层，隐含层和输出层之间的连接权值，确定极限学习机的结构

3）将以上方法应用到石化生产的过程中，验证多提出方法的可行性，并进行产量预测以及改进。

## 3.6. 工作计划进度

2016年09月~2016年10月对文献进行整理，明确课题的研究目的和意义

2016年10月~2016年04月对ELM神经网络结构进行改进，自组织确定隐含层节点数

2017年04月~2017年09月对ELM神经网络结构进行改进，优化ELM神经网络的结构

2017年09月~2017年12月在UCI数据集上验证新结构和算法的有效性并作进一步的优化

2017年12月~2018年04月完成课题，撰写论文

|  |
| --- |
| 指导教师意见：  指导教师签名：  年 月 日 |
| 审核小组意见：  审核小组组长签字：  年 月 日 |
| 研究生根据审核小组意见对开题报告的改进措施：  年 月 日 |
| 备注： |