

# 新型深度学习算法研究概述

马超, 徐瑾辉, 侯天诚, 蓝斌

(广东外语外贸大学 金融学院, 广东 广州 510006)

**摘要:** 作为一个具有巨大应用前景研究方向,深度学习无论是在算法研究,还是在实际应用(如语音识别,自然语言处理、计算机视觉)中都表现出其强大的潜力和功能.本文主要介绍这种深度学习算法,并介绍其在金融领域的领用.

**关键词:** 深度学习;受限波兹曼机;堆栈自编码神经网络;稀疏编码;特征学习

**中图分类号:** TP181 **文献标识码:** A **文章编号:** 1673-260X(2015)01-0037-03

**DOI:** 10.13398/j.cnki.issn1673-260x.2015.02.014

## 1 深度学习的研究意义

深度学习是一类新兴的多层神经网络学习算法,因其缓解了传统训练算法的局部最小性,引起机器学习领域的广泛关注.深度学习的特点是,通过一系列逻辑回归的堆栈作为运算单元,对低层数据特征进行无监督的再表示(该过程称为预学习),形成更加抽象的高层表示(属性类别或特征),以发现数据的分布式特征表示.深度学习的这种特性由于与脑神经科学理论相一致,因此被广泛应用于语音识别、自然语言处理和计算机视觉等领域.

生物学研究表明<sup>[1]</sup>,在生物神经元突触的输出变化与输入脉冲的持续时间有关,即依赖于持续一定时间的输入过程.输出信号既依赖于输入信号的空间效应和阈值作用,也依赖于时间总和效应.

传统的深度学习方法虽然较好地模拟了生物神经元的另一个重要特性——空间总和效应上的深度,却忽视了生物神经元的另一个重要特性——时间总和效应上的宽度<sup>[2]</sup>.因此,对于连续的时间变量问题(如语音识别),传统深度学习只能将连续的时间函数关系转化为空间关系,即离散化为时间序列进行处理.这样做有几个弊端:

(1)可能造成深度学习算法对时间采样频率的十分敏感,鲁棒性较差.这使得,不同时间尺度下,需要使用不同的数据和算法.这无疑是十分不方便的;

(2)导致深度网络规模过大,使得计算开销增大、学习效果变差、泛化性能降低;

(3)难以满足实际应用对算法的实时性的要求,更难以体现连续输入信息的累积效应,大大降低深度学习算法的实用性.

因此,对传统的深度学习算法进行改进,使其不但具有“深度”,亦能具有“宽度”,能够对连续时变数据进行更好的特征提取、提高算法效率和实用性,显得势在必行.基于这个

切入点,本项目借鉴时频分析与小波分析中的方法,结合数学分析领域中的泛函分析技术,与堆栈自编码神经网络相结合,提出一种新的深度学习算法——深度泛函网络.为了验证算法的有效性及其优越性,本项目将把新算法应用于金融时间序列的领域.

在目前国内外对于深度学习的研究中,几乎没有任何将深度学习技术运用于金融数据的研究.通过提出并运用得当的深度序列学习方法,我们期望从金融数据中抽取更高级的、具有经济学意义或预测性意义的高级特征(与人工设计的“技术指标”相对应),并开发相应的量化交易策略,并与其它传统算法进行对比,以说明所提算法的可行性和优越性.

## 2 国内外研究现状

人类感知系统具有的层次结构,能够提取高级感官特征来识别物体(声音),因而大大降低了视觉系统处理的数据量,并保留了物体有用的结构信息.对于要提取具有潜在复杂结构规则的自然图像、视频、语音和音乐等结构丰富数据,人脑独有的结构能够获取其本质特征<sup>[3]</sup>.受大脑结构分层次启发,神经网络研究人员一直致力于多层神经网络的研究.训练多层网络的算法以BP算法为代表,其由于局部极值、权重衰减等问题,对于多于2个隐含层的网络的训练就已较为困难<sup>[4]</sup>,这使得实际应用中多以使用单隐含层神经网络居多.

该问题由Hinton<sup>[5]</sup>所引入的逐层无监督训练方法所解决.具体地,该法对深度神经网络中的每一层贪婪地分别进行训练:当前一层被训练完毕后,下一层网络的权值通过对该层的输入(即前一层的输出)进行编码(Encoding,详见下文)而得到.当所有隐含层都训练完毕后,最后将使用有监督的方法对整个神经网络的权值再进行精确微调.在Hinton的原始论文中,逐层贪婪训练是通过受限波兹曼机

(Restricted Boltzmann Machine,RBM) 以及相对应的对比散度方法(Contrastive Divergence)完成的.与通常的神经元不同,RBM是一种概率生成模型,通常被设计为具有二元输入-输出(称为Bernoulli-Bernoulli RBM).通过对每一层的受限波兹曼机进行自底向上的堆栈(如图1),可以得到深度信念网(Deep Belief Network,DBN).

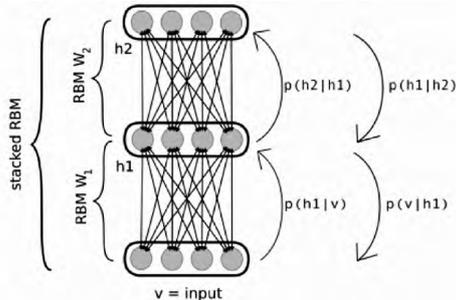


图1 堆栈 RBM 结构示意图

除了生成式的 RBM,还有其他的深度学习结构被广泛使用和研究.如堆栈自编码神经网络(Stacked Auto-Encoder Network,SAEN)<sup>[6]</sup>,以及深度卷积神经网络(Deep Convolutional Network)<sup>[7]</sup>等.前者的优势在于可以简单地采用通常的BP算法进行逐层预训练,并且引入随机化过程的抗噪声自编码网络(Denoising SAEN)泛化性能甚至超过 DBN<sup>[8]</sup>,而后者则通过权值共享结构减少了权值的数量,使图像可以直接作为输入,对平移、伸缩、倾斜等的变形具有高度不变性,因此在图像识别领域有着广泛应用.

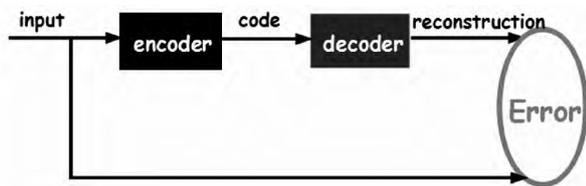


图2 自编码网络示意图

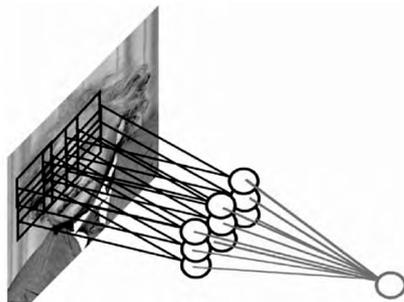


图3 卷积神经网络结构示意图.通过局部感知、权值共享以及池化,产生对图像平移、伸缩等变形的高度不变性

近年来,稀疏编码(Sparse Encoding)和特征学习(Feature Learning)成为了深度学习领域较为热门的研究方向.B.A.Olshausen<sup>[9]</sup>等针对人脑的视觉感知特性,提出稀疏编码的概念.稀疏编码算法是一种无监督学习方法,它用来寻找一组“过完备”的基向量来更高效地表示输入数据的特征,更

有效地挖掘隐含在输入数据内部的特征与模式.针对稀疏编码的求解问题,H.Lee等在2007年提出了一种高效的求解算法<sup>[10]</sup>,该算法通过迭代地求解两个不同的凸规划问题以提高效率.同年,H.Lee等发现,当训练样本为图像时,对DBN的训练进行稀疏性的约束有利于算法学习到更高级的特征<sup>[11]</sup>.例如,对手写识别数据集进行训练时,稀疏性约束下的DBN算法自主学习到了“笔画”的概念.

基于<sup>[10,11]</sup>的研究成果,R.Raina等<sup>[12]</sup>提出了“自导师学习(Self-Taught Learning)”的概念.与无监督学习(Unsupervised Learning)和半监督学习(Semi-supervised Learning)不同,自导师学习利用大量易获得的无标签数据(可以来自不同类别甚至是未知类别),通过稀疏编码算法来构建特征的高级结构,并通过支持向量机(Support Vector Machine,SVM)作为最终层分类器对少数有标签数据进行分类.这种更接近人类学习方式的模式极大提高了有标签数据的分类准确度.与之类似,H.Lee,R.Grosse等<sup>[13]</sup>提出了一种具有层次结构的特征学习算法.该算法将卷积神经网络与DBN结合,并通过稀疏正则化(Sparsity Regularization)的手段无监督地学习层次化的特征表征.图像识别实验表明,该算法能够自主学习得出“物体(Object Parts)”的概念,较好体现了人脑视觉感知的层次性和抽象性.

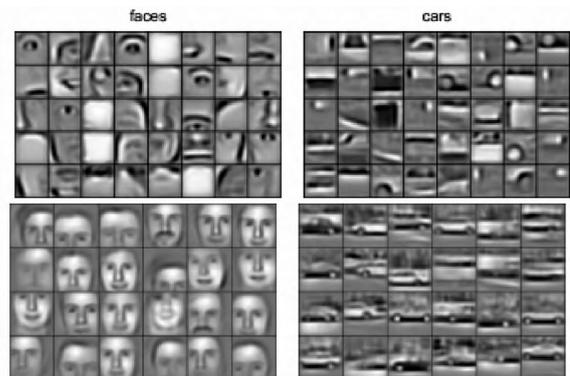


图4 层次稀疏编码在人脸识别和车辆识别中学习到的层次高级特征

### 3 发展趋势

由于信号处理、语音识别、金融时间序列分析、视频分析等领域的实时应用需求,研究能够处理连续时变变量、自然体现时间联系结构的深度学习算法(即深度序列学习,Deep Sequence Learning)成为了新的研究热点.G.W.Taylor,G.E.Hinton等<sup>[14]</sup>提出时间受限波兹曼机(Temporal RBM,TRBM).该模型使用二值隐含元和实值可视元,并且其隐含元和可视元可以与过去一段历史的可视元之间可以有向地被相连.同时,该模型被用于人类动作识别,并展现出了优秀的性能.针对TRBM的一些不足,一些改进算法也不断涌现,如<sup>[15,16]</sup>.然而,该类深度学习模型虽然考虑了动态的时间变量之间的联系,但依然只能处理离散时间问题,本质上还是属于转化为空间变量的化归法.同时,在自编码网络框架下,依

然缺乏较好解决时间过程(序列)问题的方案。

#### 4 金融时序数据中的应用

传统金融理论认为,金融市场中的证券价格满足伊藤过程,投资者无法通过对历史数据的分析获得超额利润。然而,大量实证研究却表明,中国股票价格波动具有长期记忆性,拒绝随机性假设,在各种时间尺度上都存在的可以预测的空间。因此,如何建立预测模型,对于揭示金融市场的内在规律,这无论是对于理论研究,还是对于国家的经济发展和广大投资者,都具有重要的意义。

股票市场是一个高度复杂的非线性系统,其变化既有内在的规律性,同时也受到市场、宏观经济环境,以及非经济原因等诸多因素的影响。目前国内外对证券价格进行预测的模型大致分为两类:一是以时间序列为代表的统计预测模型,该类方法具有坚实的统计学基础,但由于金融价格数据存在高噪声、波动大、高度非线性等特征,使得该类传统方法无法提供有效的工具。另一类是以神经网络、支持向量机等模型为代表的数据挖掘模型。该类模型能够处理高度非线性的数据,基本上从拟合的角度建模。虽然拟合精度较高,但拟合精度的微小误差往往和市场波动互相抵消,导致无法捕捉获利空间甚至导致损失,外推预测效果无法令人满意。因此,建立即能够处理非线性价格数据,又有良好泛化能力的预测模型势在必行。

#### 参考文献:

- [1]Zhang L I, Tao H W, Holt C E, et al. A critical window for cooperation and competition among developing retinotectal synapses[J]. Nature, 1998, 395(6697).
- [2]37-44. 欧阳楷, 邹睿. 基于生物的神经网络的理论框架——神经元模型[J].北京生物医学工程,1997,16(2):93-101.
- [3]Rossi A F, Desimone R, Ungerleider L G. Contextual modulation in primary visual cortex of macaques [J]. the Journal of Neuroscience, 2001, 21(5): 1698-1709.
- [4]Bengio Y. Learning deep architectures for AI [J]. Foundations and trends? in Machine Learning, 2009, 2(1):1-127.
- [5]Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18(7): 1527-1554.
- [6]Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the 25th international conference on Machine learning. ACM, 2008: 1096-1103.
- [7]Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009: 609-616.
- [8]Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion [J]. The Journal of Machine Learning Research, 2010, 9999: 3371-3408.
- [9]Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: A strategy employed by V1?[J]. Vision research, 1997, 37(23): 3311-3325.
- [10]Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms [J]. Advances in neural information processing systems, 2007, 19: 801.
- [11]Lee H, Ekanadham C, Ng A Y. Sparse deep belief net model for visual area V2 [C]//NIPS. 2007, 7: 873-880.
- [12]Raina R, Battle A, Lee H, et al. Self-taught learning: transfer learning from unlabeled data[C]//Proceedings of the 24th international conference on Machine learning. ACM, 2007: 759-766.
- [13]Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009: 609-616.
- [14]Taylor G W, Hinton G E, Roweis S T. Modeling human motion using binary latent variables[J]. Advances in neural information processing systems, 2007, 19: 1345.
- [15]Sutskever I, Hinton G E, Taylor G W. The Recurrent Temporal Restricted Boltzmann Machine [C]//NIPS. 2008, 21: 2008.
- [16]Lockett A J, Miiikkulainen R. Temporal convolution machines for sequence learning [J]. To Appear, 2009: 1-8.