

支持向量机和神经网络的融合发展

李应红, 尉询楷

(空军工程大学工程学院, 陕西西安 710038)

摘要:提出了支持向量机和神经网络的融合发展观。分析了支持向量机和神经网络的异同点。从认知模型角度探讨了神经网络认知模型对于支持向量机认知模型发展的指导作用,提出了支持向量机认知模型概念和发展思路;从支持向量机算法思想角度,提出了一类神经网络算法的发展。

关键词:机器学习;支持向量机;神经网络;融合发展

中图分类号: TP18 **文献标识码:** A **文章编号:** 1009-3516(2005)04-0070-04

传统机器学习方法的经验风险最小化原则隐含地使用了训练样本无穷多的假设条件。然而实际应用中,这种前提条件尤其当问题处于高维空间时往往得不到满足。为解决这个问题,Vladimir N. Vapnik 及其同事早在 20 世纪 60 年代就开始研究有限样本情况下的机器学习问题,直到 90 年代中期,形成了一个较完善的有限样本理论体系——统计学习理论^[1],并发明了支持向量机。统计学习理论提出了结构风险最小化原则,该原则使得学习机器在可容许的经验风险范围内,总是采用具有最低复杂度的函数集。该理论致力于寻找在小样本情况下学习问题的最优解,使得统计学习理论在小样本下同样能具有推广价值。

统计学习理论已经成为研究热点。近些年来出现了很多发展和改进的支持向量机算法并已经成功应用到了手写体识别,人脸识别与检测,文本分类,时间序列分析,信号处理,基因、蛋白质分析,预测,控制,语音识别,图像分类,故障诊断等众多研究领域。

1 支持向量机和神经网络的比较

神经网络和支持向量机产生背景不同,导致两种机器学习方法有着较大的差异。二者的异同是研究者,特别是工程技术人员在应用时非常关心的问题。

1.1 支持向量机与神经网络对于输入数据处理的差异

当训练样本维数很高时,神经网络训练和测试结果往往会出现较大的随机性。究其原因,首先,神经网络的解往往收敛于局部极值点,而在高维空间中可能存在众多的局部极值点,局部极值点之间有很大的差异,因此神经网络的解必然呈现出随机性。其次,高维问题与小样本数是紧密联系的。样本数的多少是相对的,在低维空间较少样本数就可以比较充分描述整个样本空间,而在高维空间,同样程度下,需要的样本数就会随着维数的增加以指数形式增长,训练样本数与输入空间维数之间一个粗略的估计式为:样本数量 $M \approx C\varepsilon^n$, C 为常数, ε 为要求的逼近误差。另外,在运算方面,当特征空间的维数增加时,神经网络运算量急剧增加,从而使训练和测试时间过长。因此,应用神经网络时需要对数据实行降维操作。

上面描述的问题实际就是维数灾难,支持向量机通过引入核函数有效解决了这一问题。核函数的引入使得操作可以直接在输入空间进行而不必在潜在的高维特征空间进行,而根据再造核 Hilbert 空间理论,在特征空间的内积总可以在输入空间找到一个相对应的核,因此,支持向量机可以方便的处理高维数据问题,而不需要事先对高维数据降维。

1.2 支持向量机与神经网络最优解的差异

收稿日期:2004-11-16

基金项目:总装预研基金资助项目(41327030103)

作者简介:李应红(1963-),男,重庆奉节人,教授,博士生导师,主要从事航空发动机等研究。

神经网络主要的优化方法是梯度修正算法,而梯度修正算法不能收敛到全局最优点。当然,有的研究者设计了可以跳出局部极值的算法(例如与遗传算法、混沌理论、模拟退火等结合使用),在实际应用中取得了一定的效果,但是仍然不能保证一定能寻到全局最优解。另外,神经网络学习过程的参数选择也是一件困难的事情。如果参数选择不当,则可能导致网络不收敛,或者收敛速度慢得难以接受。为了达到较好的训练效果,需要多次尝试、比较。与之相比,支持向量机通过求解凸二次规划获得全局的最优解,由于采用了特殊的损失函数使得求解的模型具有一定的稀疏性,可有效降低模型复杂性,这是神经网络所不能比拟的。神经网络的求解依赖于所有的数据,几乎对所有数据都有依赖性。因而,同样情况下神经网络结构复杂性偏大,虽可有较好的内插特性,但外推则可能由于不相关点的影响使预测结果较差。而支持向量机仅取决于数据中的关键点——支持向量,且可有效控制网络容量,因而对于未知数据点具有较好的预测性能。

1.3 支持向量机与神经网络功能结构上的差异

结构选择是神经网络应用中的另一个难点问题。如何选择网络结构和参数,目前没有统一的原则,而需由问题本身和研究者的经验确定。样本数一定情况下,隐层结点越多,则网络越复杂,其学习能力越强,而推广能力变差,容易导致过学习现象;反之,如果隐层结点数太少,则其学习能力不足,难以有效地从数据中学习,而产生欠学习问题。相比较而言,支持向量机依据算法的先进性使得计算单元数目能够在求解凸二次规划之后最优确定,而且这些计算单元具有有限可控的学习能力。从这方面来看,神经网络的计算单元数目则是需要事先确定好的,而且每个单元的学习能力是不可控的,因而如果神经网络设计过程中,停机准则选择不恰当,就会出现过学习使得网络性能变差甚至不具有推广性。

1.4 支持向量机的缺点与神经网络的优点

任何事物都不是完美的。一定意义上讲,支持向量机是将确定神经网络隐层节点个数的难题转化成为核函数的选取问题,研究表明不同的核函数往往具有不同的分类结果。支持向量机对于核函数的选取和构造也是没有解决的难题。输入数据集的大小和结构风险界仍然是支持向量机十分关心的问题,当输入数据集很大时,此时的核矩阵将会很大,计算的速度也会下降较大。另外,对于风险界的最小化并不一定保证得到的结果具有最好的推广性,因此一个好的模型也需要多次实验验证后才能确定。

支持向量机对于多元问题的处理上至今仍没有很好的解决办法,对于分类问题支持向量机将多类转化为二类的组合,对于多元回归问题,支持向量机只能采用适当的方法构造多个单值回归模型。而神经网络则可以方便的处理多元问题,其特有的网络结构可以方便地容纳各种先验知识,使得设计的神经网络能够更好的符合实际问题的需要。

2 神经网络认知模型和支持向量机认知模型

模式分类过程总是存在对分类器的训练和测试过程,当中的训练过程就好比是人类对于万事万物的认识过程,也就是“认”。而对于测试过程好比是人类的识别过程,也就是“知”。总的来看,模式分类的过程就如同人类的认知过程一样。事实上,人类认识事物和传统的模式分类识别过程是有区别的。人类侧重于认识,只有在细小之处才重视区别,而传统模式识别则只注意“区分”,没有重视“认识”的概念。中科院王守觉院士结合人类对事物的认知过程和特点,对神经网络的多维空间几何分析及其理论^[2]进行了较为深入的研究,在此基础上提出把模式识别问题看成是模式的“认识”,而不是模式分类,是对一类样本的“认识”。为了强调与传统模式识别在概念上的不同,王守觉院士采用了“仿生模式识别”这一概念^[3],仿生模式识别是从“认识”模式的角度进行模式识别,更接近于人类的识别。与神经网络相对应,我们提出了支持向量机认知模型的概念,在现有的文献中,我们发现一些文献与仿生模式识别具有类似的思想。为了区别起见,统一将其称为支持向量机简单认知模型^[4]。

Tax 等人提出了一种新的数据描述算法——支持向量域描述(Support Vector Domain Description, SVDD)^[5],主要用来进行数据描述和剔除奇异点。SVDD 的主要思想是通过计算包含一组数据的最小超球形边界来对该组数据进行描述。在进行模式识别时,只需要对已知类别的数据进行描述,得到描述边界,根据此边界就可以判断未知的数据是否属于该类别。

陆从德等人提出了一种基于支持向量域描述的学习分类器^[6]。在两类样本分类中,该算法在训练时通过对已知类别样本的描述求取包含该类样本的球形边界,然后通过该边界对两类样本数据进行分类,在求取

边界的优化问题中,采用乘性规则来直接求取拉格朗日乘子。这种方法尤其适用于具有已知正类样本和少数负类样本的情形。

朱美琳等人提出用球结构的支持向量机解决多分类问题^[7],其主要思想就是将同一类数据用超球来界定,数据空间变为由若干个超球组成,在三维上面象是很多肥皂泡的集合。球分类方法对于新增加的分类只需对其本身进行一个二次规划以及其它简单计算就可以建立新的分类体系,方法易于操作,具备扩展能力。

应当指出仿生模式识别的概念对于促进模式识别领域的发展具有较大的意义,但是仿生模式识别中采用的神经网络方法本质上仍存在一些弊端。笔者认为如何在仿生模式识别的概念下采用新形式的替代工具是发展的重点。支持向量机简单认知模型是对仿生模式识别的一种新发展。在实际模式识别的应用中,更希望能够做到任意形状覆盖。笔者认为有两种思路可以发展。一是发展结合多球体的空间覆盖方法,利用一组 SVDD 对同一类别进行数据描述,将其获得的联合覆盖区域作为整个类别的边界,以提高现有单一球体覆盖精度差的问题;二是发展新形式的支持向量数据描述方法,可否发展一种能够进行任意形状覆盖的新型数据描述方法。笔者认为对于第一种思路实现相对简单,但是算法复杂性会明显增高,决策过程也相对较为烦杂。笔者认为将二种思路与计算几何的形体构造理论相结合、借鉴先进的计算几何的思想或许会产生新型的数据描述方法,而且发展新形式的数据描述方法有助于发展、完善支持向量机的理论,有助于推动模式识别领域的发展,是一种较为实用的发展思路。

3 支持向量机理论指导神经网络的发展

神经网络在经过 20 世纪 80 年代热潮之后,理论发展缓慢,很多研究人员退出了这一领域,主要原因是神经网络理论上较多的缺陷无法圆满解决,例如结构选择、局部极值、过学习等问题。虽然围绕这些问题都有一定的解决办法,但大多不能从根本上解决神经网络的弊端。

随着统计学习理论的深入发展,其理论上的先进性给神经网络的发展提供了启示。神经网络理论的研究人员认识到如何保证网络的推广性,如何设计网络使它在有限(少量)的样本下学习后,具有好的推广性才是神经网络理论和工程应用中亟需解决的关键问题。

笔者在支持向量机最小化经验风险、最小化推广误差上界的思想指引下,提出了一类改进的机器学习算法^[8],该算法兼顾神经网络的学习能力和推广特性,采用随机数任意选取输入权值而将单隐层神经网络转化成为线性系统,通过引入矩阵的极小范数最小二乘解获得最佳的网络输出参数,网络可以保证获得最小误差的同时具有良好的推广性。

本文提出了支持向量机和神经网络的融合发展观,对于机器学习领域更好、更快的发展具有指导意义。支持向量机理论正处于高速发展的阶段,然而如何减少理论发展、应用中的盲目性,增强理论研究的针对性和有效性是研究者应当引起高度重视的问题。目前,神经网络的发展趋向于理智、正常的发展阶段,而人们对于支持向量机理论的研究工作则带有一定热情和盲从。借鉴神经网络理论的发展历史和神经网络理论的优点,融合支持向量机理论的先进性,客观、理智的推进支持向量机和神经网络共同发展是作者撰写本文的目的。

参考文献:

- [1] Vapnik V N. 统计理论学习的本质[M]. 张学工. 北京:清华大学出版社,2000.
- [2] 王守觉,王柏南. 人工神经网络的多维空间几何分析及其理论[J]. 电子学报,2002,30(1):1-4.
- [3] 王守觉. 仿生模式识别(拓扑模式识别)——一种模式识别新模型的理论与应用[J]. 电子学报,2002,30:1417-1420.
- [4] 李应红,尉询楷,刘建勋. 支持向量机的简单认知模型及拓扑模型分析[R]. 西安:空军工程大学工程学院,2004.
- [5] Tax D, Duin R. Support Vector Domain Description[J]. Pattern Recognition Letters, 1999, 20:1191-1199.
- [6] 陆从德,张太镒,胡金燕. 基于乘性规则的支持向量域分类器[J]. 计算机学报,2004,27(5):690-694.
- [7] Zhu Meilin, Wang Yue, Chen Shifu. Sphere - Structured Support Vector Machines for Multi - Class Pattern Recognition[A]. In: RSFDGrC 2003, LNAI 2639[C],2003.
- [8] 尉询楷,李应红. 一种新的机器学习算法及其与支持向量机的比较[A]. 第 14 届中国神经网络会议,合肥:2004.

(编辑:姚树峰)

Fusion Development of Support Vector Machines and Neural Networks

LI Ying - hong, WEI Xun - kai

(The Engineering Institute, Air Force Engineering University, Xitm, Shaanxi 710038, China)

Abstract: This paper proposes a standpoint of fusion development of support vector machines (SVM) and neural networks (NN), and argues that SVM and NN are equivalent in function but different in implementation algorithm.

In term of cognitive model, it discusses the guidance effect of NN cognitive model on SVM counterpart development, and simultaneously investigates the concepts and development methods of SVM cognitive model in detail. Finally, it proposes a class of NN algorithm development from the angle of core idea of SVM algorithm.

Key words: machine learning; support vector machines; neural networks; fusion development

(上接第 54 页)

参考文献:

- [1] Wu Y, Woerner B D. The Influence of Quantization and Fixed Point Arithmetic upon the BER Performance of Turbo Codes [A]. In: Pro VTC99 Fall [C]. 1999.
- [2] Michel H, When N. Turbo - Decoder Quantization for UMTS [J]. IEEE Communication Letters, 2001, 5: 55 - 57.
- [3] 孙 蓉, 陈 军. Turbo 码 MAP 译码算法中量化问题的研究 [J]. 电子学报, 2001, 29(10): 1536 - 1539.
- [4] Robertson P, Hoeher P. Optimal and Sub - optimal Maximum a Posteriori Algorithm Suitable for Turbo Decoding [J]. European Trans on Telecomm, 1997, 8: 119 - 125.
- [5] 闫 涛, 杜兴民. 外信息在 Turbo 迭代译码中的使用研究 [J]. 空军工程大学学报(自然科学版), 2004, 5(5): 72 - 76.
- [6] Jordan M, Nichols R. The Effects of Channel Characteristics on Turbo Code Performance [A]. In: Proc Milcom96 [C]. 1996.
- [7] Hekstra A. An Alternative to Metric Rescaling in Viterbi Decoders [J]. IEEE Trans on Communication, 1989, 37(11): 1220 - 1222.
- [8] Worm A, Michel H. Advanced Implementation Issues of Turbo Decoders [A]. In: Proc 2nd int Symp on Turbo Codes [C]. 2000.
- [9] Masera G, Piccinini G. VLSI Architectures for Turbo Codes [J]. IEEE Trans on VLSI Systems, 1999, 7(3): 369 - 378.
- [10] Sharma S, Attri S. A Simplified and Efficient Implementation of FPGA - Based Turbo Decoder [A]. In: Proc IEEE Conference on Performance, Computing and Communications [C]. 2003.

(编辑: 姚树峰)

Analysis and Study of Quantization of Turbo Codes Based on Log -MAP Algorithm

YAN Tao, DU Xing - min

(The Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: For an AGN channel, based on Log - MAP decoding algorithm, quantization and fixed - point number representation of all the data which are produced during the SISO iterative decoding of turbo codes are analyzed and studied systematically. According to data observation and anatomy of algorithm, with Monte Carlo simulation, this paper presents the fixed - point representations and quantization details of the data such as received information, extrinsic information, the channel reliability value, state metric, etc. By the quantization scheme determined in this paper a better tradeoff between the decoding performance and implementation complexity can be obtained, and the degradation of performance is not more than 0.15 dB.

Key words : Log - MAP algorithm, quantization ; received data ; reliability of channel ; extrinsic information ; state metric