

一种深度学习的信息文本分类算法

吕淑宝, 王明月, 翟祥, 陈宇

(东北林业大学 信息与计算机工程学院 黑龙江 哈尔滨 150040)

摘要: 针对传统文本分类算法准确率低和正确率分布不均匀的问题,提出了基于深度学习的文本分类算法。深度信念网络具有强大的学习能力,可以从高维的原始特征中提取高度可区分的低维特征,不仅能够更全面的考虑到文本信息量,而且能够进行快速分类。采用 TF-IDF 方法计算文本特征值,利用深度信念网络构造分类器进行精准分类。实验结果表明,与支持向量机、神经网络和极端学习机等常用分类算法相比,该算法有更高的准确率和实用性,为文本的分类研究开拓了新思路。

关键词: 文本分类; 深度信念网络; 分类器

DOI: 10.15938/j.jhust.2017.02.020

中图分类号: TP181 **文献标志码:** A **文章编号:** 1007-2683(2017)02-0105-07

AN Information Text Classification Algorithm Based on DBN

LÜ Shu-bao, WANG Ming-yue, ZHAI Xiang, CHEN Yu

(School of Information and Computer Science, Northeast Forestry University, Harbin 150040, China)

Abstract: Aiming at the problem of low categorization accuracy and uneven distribution of the traditional text classification algorithms, a text classification algorithm based on deep learning has been put forward. Deep belief networks have very strong feature learning ability, which can be extracted from the high dimension of the original feature, so that the text classification can not only be considered, but also can be used to train classification model. The formula of TF-IDF is used to compute text eigenvalues, and the deep belief networks are used to construct the classifier. The experimental results show that compared with the commonly used classification algorithms such as support vector machine, neural network and extreme learning machine, the algorithm has higher accuracy and practicability, and it has opened up new ideas for the research of text classification.

Keywords: text classification; deep belief network; classifier

收稿日期: 2015-12-07

基金项目: 中央高校基本科研业务费专项资金(2572015DY07); 黑龙江省自然科学基金(F201347); 哈尔滨市科技创新人才专项资金(2013RFQXJ100); 国家自然科学基金(61300098); 教育部大学生创新训练计划项目(201510225043)。

作者简介: 吕淑宝(1991—),男,硕士研究生;

陈宇(1975—),男,博士后,副教授,硕士研究生导师。

通信作者: 王明月(1991—),女,硕士研究生。

0 引言

文本分类是根据提前建立好的分类器^[1-2],让计算机对给定的未知类别的文本集进行分类的过程^[3]。一个完整的文本分类过程主要包括以下几个部分:首先是文本预处理,将文本表示成易于计算机处理的形式;其次是文本向量表示;再次是根据训练集(具有类标签)学习建模,构建出分类器;最后利用测试集测试建立好的分类器的性能,并不断反馈、学习提高该分类器性能,直至达到预定的目标。

常用的文本分类算法包括 SVM (support vector machine, 支持向量机)^[4], BP (back propagation) 神经网络^[5-6]、ELM (extreme learning machine, 极端学习机) 算法^[4]等^[7], 这些方法往往存在着有限样本或局部最优及过学习问题^[2], 并且这些方法为了避免维度灾难通常对数据做降维处理, 此时得到的结果并不能很好的反应文本的特征, 最终导致分类准确率低下。因此, 本文主要研究如何文本分类的准确率和效率。

目前国内对于文本分类的研究比较少, 因此如何提高文本分类的准确率, 是一项重要的研究内容^[8]。

深度学习是通过构建具有多个隐层的机器学习模型组合低层特征形成更加抽象的高层特征来表示属性类别, 以发现数据的分布式特征^[9]。深度学习具有很强大的自我学习挖掘特征的能力, 可以得到最接近数据本质的表达和模式, 能够极大的提高预测、分类等性能^[10]。2006年, Geoffrey Hinton 提出深度信念网络 (deep belief network, DBN)^[11]。它是一种概率生成模型, 通过训练其神经元间的权重可以让整个网络按照最大概率来生成训练数据, 从而实现特征识别和数据分类。

深度信念网络由多层神经网络构成, 且这些结构之间均是非线性映射的, 这使得它可以很好的完成复杂函数的逼近, 从而实现快速的训练。此外, 深度信念网络输入数据的重要驱动变量是通过贪心算法逐层训练每一层网络得到的, 学习过程中既保证了数据本质特征的抽取, 同时也避免了过拟合现象的出现。

基于上述原因, 提出了基于深度学习的文本分类算法。首先, 利用中科院中文分词系统对文本进行预处理; 其次, 使用 TF-IDF (Term Frequency - Inverse Document Frequency) 公式计算文本分词的特

征值, 构成初始特征矩阵^[12]; 然后, 利用 DBN 良好的特征学习能力, 从原始的高维特征中自动提取出高度可区分的低维特征; 最后, 将得到的特征输入 softmax 回归分类器实现分类。

1 关键技术

1.1 文本的获取

手动建立文本库, 在网络上大量搜集资料汇总, 总结出 5 类文本: 花、树木、虫、土壤、水类文本, 同样本文研究的算法也适用于其他类型的文本, 本文选取的实验样本如图 1~3 所示 (经过分词后的文本)^[13]。

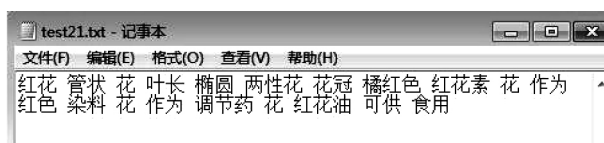


图1 花类样本

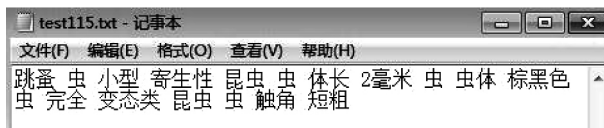


图2 虫类样本

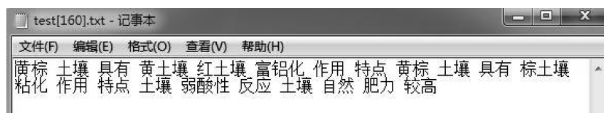


图3 土壤类样本

从所获得的数据可知, 不同类别的样本含有一些可将该类文本与其他类别文本区分开的特征词。由图 2 可看出花类文本中, 典型特征词有花、花蕊、花柄等专业性术语^[13]; 由图 2 可看出虫类文本中, 含有的关键特征词包括虫、触角、尾须等。由图 3 可看出土壤类文本中, 含有的关键特征词包括土、土壤、肥力等。树木类、水类文本含有的典型特征词显然也同样如此。

通过 ICTCLAS 系统 (institute of computing technology, chinese lexical analysis system) 对初始文本进行分词和去停用词, 从而得到实验所需的文本特征词, 然后使用经典的 TF-IDF 公式计算特征词的权值, 构成文本初始特征矩阵进行实验^[13]。

1.2 文本表示

假设所有的文本共有 n 个特征, 形成 n 维的向

量空间,每一个文本 d 可用 n 维的特征向量来表示:

$$V(d) = (T_1, W_1(d); T_2, W_2(d); \dots; T_n, W_n(d)) \quad (1)$$

式中: T_i 为文本的分词; $W_i(d)$ 为 T_i 在文本 dd 中的权值,通过 TF-IDF 公式计算文本分词的权值^[14]:

$$\begin{aligned} w_i(d) &= \frac{TF(t_i) \times IDF(t_i)}{\sqrt{\sum_{i=1}^n (TF(t_i) \times IDF(t_i))^2}} \\ &= \frac{TF(t_i) \times \log(\frac{N}{n_i} + L)}{\sqrt{\sum_{i=1}^n (TF(t_i) \times \log(\frac{N}{n_i} + L))^2}} \quad (2) \end{aligned}$$

式中: $w_i(d)$ 为特征词 T_i 的权值; $TF(t_i)$ 为特征词 T_i 在 d 中出现的频率; N 代表样本总个数; n_i 为出现 T_i 的样本的个数。

1.3 文本分类算法

传统的支持向量机算法、BP 神经网络算法等受样本影响较大,因此不同样本分类的正确率差异较大,基于深度学习的文本分类算法实验得到了较好的效果,分类结果精准并且正确率分布均匀^[8]。

使用深度信念网络进行文本分类主要包括两个过程:DBN 预训练和网络调优。其他的分类算法为了避免维度灾难往往需要对数据进行降维处理,DBN 可以自动从高维的原始特征中提取高度可区分的低维特征,因此不需要对数据进行降维就可以直接开始训练分类模型,同时也考虑了文本足够的信息量。在调优 DBN 的过程中,可以利用 DBN 的权值初始化 BP 神经网络的各层的权值,而不是利用随机初始值初始化网络,将 DBN 展开成 BP 神经网络,最后利用 BP 算法微调整个网络的参数,从而使得网络的分类性能更优,这个措施使 DBN 克服了 BP 网络因随机初始化权值参数而容易陷入局部最优的缺点。

1.3.1 DBN 预训练

深度信念网络是一种深层非线性的网络结构,它通过构建具有多个隐层的模型来组合低层特征形成更加抽象的高层特征^[9]来挖掘数据特征实现文本分类。假设 S 是一个系统,它包含有 n 层(S_1, S_2, \dots, S_n),若输入用 I 表示,输出用 O 表示,则可表示为: $I \geq S_1 \geq S_2 \geq \dots \geq S_n \geq O$,不断调整系统中的参数,使得系统的输出仍然是输入 I ,那么我们就可以自动得到输入 I 的层次特征,即 S_1, S_2, \dots, S_n 。DBN(深度信念网络)是一个概率生成模型,它建立了观测数据和标签之间的联合分布^[15]。

DBN 是由一层层的 RBM(restricted boltzmann machine,受限玻尔兹曼机)不断堆叠而成的。RBM 是一种典型的神经网络,如图 4 所示。

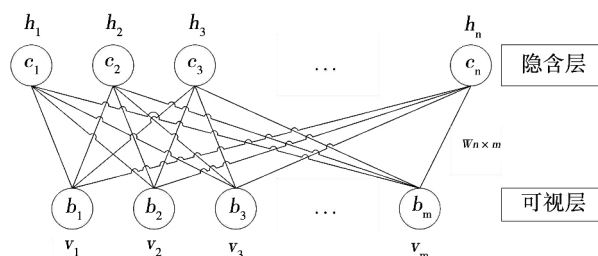


图4 RBM 的神经结构

RBM 网络共有 2 层,其中第一层称为可视层,一般来说是输入层,另一层是隐含层,也就是我们一般指的特征提取层。该网络可视层和隐层中神经元彼此互联,隐层单元被训练去捕捉可视层单元所表现出来的数据的高阶相关性^[2]。图 4 中 $w_{n \times m}$ 为可视层与隐层之间的权重矩阵; b 为可视层节点的偏移量; c 为隐层节点的偏移量; v 为可视层节点的状态向量; h 为隐层节点的状态向量。

在训练 DBN 时,采用贪婪法逐层训练每一层的 RBM。前一层的 RBM 训练完成后,将其结果作为下一层 RBM 的输入来训练该层 RBM,以此类推训练完若干 RBM,从而构建完整的 DBN 网络,其网络结构如下图 5 所示^[2]。

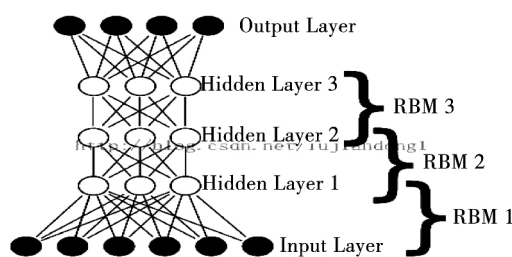


图5 DBN 的网络结构

RBM(受限玻尔兹曼机)是一种基于能量的模型,如图 4 中所示的 RBM 中,可视层变量 v 和隐层变量 h 的联合配置的能量为

$$E(v, h | \theta) = \frac{1}{2} (v^T w h + b^T v + c^T h) \quad (3)$$

其中 $\theta = (w, b, c)$ 为参数集合,当参数确定后, RBM 的归一化因子为 $Z(\theta) = \sum_{v, h} e^{-E(v, h | \theta)}$,此时基于能量函数,我们可以得到 (v, h) 的联合概率分布:

$$p(v, h | \theta) = \frac{e^{-E(v, h | \theta)}}{Z(\theta)} \quad (4)$$

隐层节点的条件概率如下:

$$p(h_j = 1 | v) = \delta(c_j + \sum_i v_i w_{ij}) \quad (5)$$

可视层节点的条件概率如下:

$$p(v_i = 1 | h) = \delta(b_i + \sum_j w_{ij} h_j) \quad (6)$$

RBM 的训练过程, 实际上是求出一个最能产生训练样本的概率分布。也就是说, 要求一个分布, 在这个分布里, 训练样本的概率最大。由于这个分布的决定性因素在于权值 w , 所以我们训练 RBM 的目标就是寻找最佳的权值。

Hinton^[16] 提出了一种快速算法, 称作对比分歧 (contrastive divergence, 简称 CD 算法) 算法。这种算法只需迭代 k 次, 就可以获得对模型的估计, 而 k 通常等于 1。CD 算法开始是用训练数据去初始化可视层, 然后用条件分布计算隐层; 然后, 再根据隐层用条件分布来计算可视层。这样产生的结果就是对输入的一个重构。具体来说, 在这个训练阶段, 在可视层会产生一个向量 v , 通过它将值传递到隐层。反过来, 可视层的输入会被随机的选择, 以尝试去重构原始的输入信号。最后, 这些新的可视的神经激活单元将前向传递重构隐层激活单元, 获得 h 。这些步骤就是我们熟悉的 Gibbs 采样, 权值更新的主要依据是隐层激活单元和可视层输入之间的相关性差别。根据 CD 算法:

$$\Delta w_{ij} = \varepsilon (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}) \quad (7)$$

其中: ε 代表学习率, $\langle v_i h_j \rangle_{\text{data}}$ 指的是样本数据的期望, $\langle v_i h_j \rangle_{\text{recon}}$ 是重构后可视层数据的期望。

DBN 的预训练过程:

- 1) 用贪婪算法训练第一个 RBM;
- 2) 固定第一个 RBM 的权值和偏置值, 将其结果的输出作为较上一层 RBM 的输入;
- 3) 重复以上步骤任意多次, 直到重构误差尽可能的小, 此时隐藏层可作为可视层输入数据的特征;

DBN 预训练算法具体步骤如下:

输入: 训练样本 x_0 , 可视层与隐藏层单元个数 n, m , 学习率 ε , 最大训练周期 T ;

输出: 权重矩阵 w , 可视层偏置量 a , 隐藏层偏置量 b ;

步骤 1: 初始化可视层单元初始状态 $v_1 = x_0$, W, a, b 为随机较小的数字;

步骤 2: 迭代训练周期 t ;

步骤 3: 通过可视层 v_1 计算隐层 h_1 , 具体为循环计算 $P(h_{1j} = 1 | v_1)$ 值, 并以该概率值为隐藏层第 j 个单元取值为 1 的概率;

步骤 4: 通过隐藏层 h_1 计算可视层 v_2 , 具体为

循环计算 $P(v_{2i} = 1 | h_1)$ 值, 并以该概率值为可视层第 i 个单元取值为 1 的概率;

步骤 5: 通过可视层 v_2 计算可视层 h_2 , 具体为循环计算 $P(h_{2j} = 1 | v_2)$ 值, 并以该概率值为可视层第 j 个单元取值为 1 的概率;

步骤 6: 更新参数 W, a, b

$$W \leftarrow W + \varepsilon (P(h_1 = 1 | v_1) v_1^T - P(h_2 = 1 | v_2) v_2^T) \quad (8)$$

$$a \leftarrow a + \varepsilon (v_1 + v_2) \quad (9)$$

$$b \leftarrow b + \varepsilon (P(h_1 = 1 | v_1) - P(h_2 = 1 | v_2)) \quad (10)$$

步骤 7: 判断是否达到迭代次数, 是转到步骤 8, 否转到步骤 2;

步骤 8: 输出参数 w, a, b 结束;

1.3.2 网络调优

在确定了网络的各个权值后, 再次根据样本, 以 BP 神经网络的算法, 进行一次有监督的学习过程, 这一过程被称为深度信念网络的调优。如下图 6 所示。

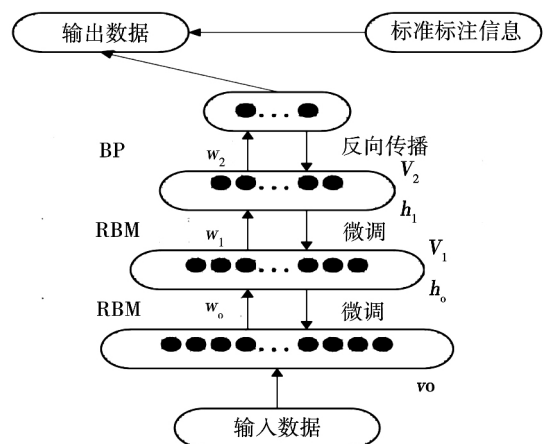


图6 网络调优

在 DBN 的最后一层设置 BP 网络, 接收 RBM 的输出特征向量作为它的输入特征向量, 有监督地训练分类器, 然而每一层 RBM 只能确保自身层内的权值对该层特征向量映射达到最优, 并不是对整个 DBN 的特征向量映射达到最优。BP 算法可以利用反向传播网络将错误信息自顶向下传播至每一层 RBM 来调整整个 DBN 网络, 提高该网络的分类性能。

2 实验结果

实验首先针对文本进行样本选择, 如表 1 所示。

表 1 实验样本选取数量表

	花类	树木类	虫类	土壤类	水类
训练	1 200	1 200	1 200	1 200	1 200
测试	200	200	200	200	200

表 1 所示 ,实验选取 5 个类别:花、树木、虫、土壤、水。训练样本每个类别选取1 200 个 ,共6 000 个。对训练样本加高斯白噪声处理得到测试样本 ,每个类别选取 200 个 ,共 1 000 个测试样本 ,样本总数 7 000 个。

对样本预处理之后使用 TF-IDF 方法计算文本特征词的权值 ,构成训练样本的初始特征值矩阵 ,然后对训练样本的初始特征值矩阵加高斯白噪声获得测试样本初始特征值矩阵 ,最后进行分类操作。如图 7 所示为基于深度学习的文本分类的流程图。

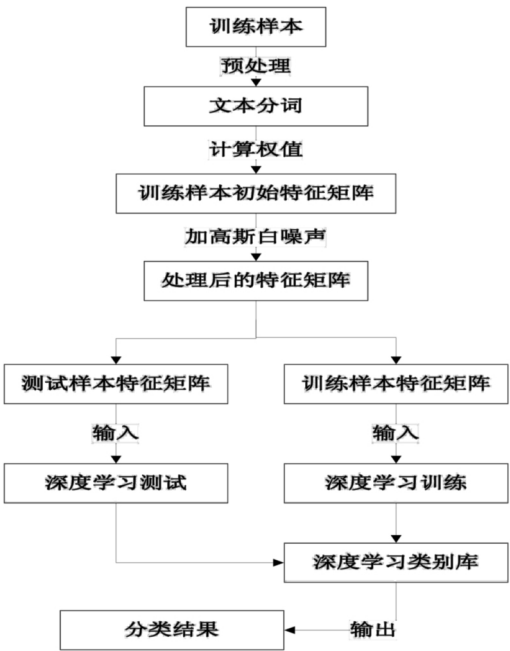


图 7 构造深度信念网络分类器的流程图

实验过程中 ,训练和测试样本初始特征矩阵维数均为1 127 维 ,文本类别总数 $K=5$ 。初始化参数 ,DBN 网络的节点数分别为 1127 - 1700 - 1000 - 25 - 5 ,共 5 层。从图 8 中可以看出随着深度学习迭代次数的增加 ,目标函数的值(也即重构方差的值)在不断地递减 ,逐渐接近最优值 0.1。此外 ,在更新过程中 ,我们可以看到 err (重构的方差)明显在递减 ,最开始有 40 000 多 (总共 1 127 维 ,可见最开始重构非常差) ,但最后只有 0.1 的 err ,说明最后重构已经比较准确 ,非常接近原始数据的分布状态。

表 2 为 DBN 算法的迭代次数所对应的重构方差数据。由表 2 可见 ,在隐层节点数一定时 ,随着迭

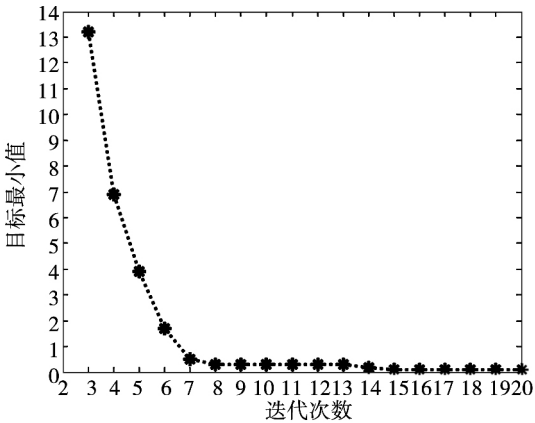


图 8 DBN 在文本数据集上的收敛图

代次数的增加 ,可以降低算法的误差 ,但是迭代次数越多 ,训练时间也越久。在隐含层节点数为 1700 - 1000 - 25 时 ,迭代次数为 20 时 ,DBN 达到重构方差最小值 0.1 ,并且趋于稳定。所以此时训练时间最短 ,分类性能最好。

表 2 迭代次数对算法的影响

隐层节点	迭代次数				
	10	20	30	40	50
1200 - 300 - 150	12.7	0.8	0.7	0.6	1.1
1000 - 300 - 150	8.3	1.3	1.3	1.7	1.8
1700 - 1000 - 150	2.8	0.2	0.3	0.4	0.6
1700 - 1100 - 150	0.9	0.2	0.3	0.5	0.7
1700 - 1000 - 500	1.4	1.3	2.0	2.5	2.9
1700 - 1000 - 100	0.3	0.2	0.2	0.3	0.4
1700 - 1000 - 25	0.6	0.1	0.1	0.1	0.1

表 3 为实验的训练时间和实验误差对比 ,由表 3 可见 ,DBN 与极端学习机算法的测试误差率较低 ,分类结果比较理想 ,但是 DBN 优于 ELM ,它以牺牲时间为代价提高了分类的准确率。

表 3 4 种算法的性能对比图

算法	训练时间	测试误差
SVM	2.30	0.236
BP	735.48	0.216
ELM	0.249 6	0.192
DBN	4 732.79	0

为了对比基于深度信念网络的分类效果 ,我们选取相同数量的样本 ,每类样本各 200 个 ,总共 1 000 个测试样本 ,为了避免维度灾难 ,使用 PCA (主成分分析)的方法处理初始特征矩阵 ,得到降维后的特征矩阵 ,然后再用 SVM、BP、ELM、这三种方法

进行分类。其中 BP 采用是 3 层的网络结构,节点数分别为 213-12-5,迭代 200 次。使用 libsvm-3.11 工具箱,进行 SVM 实验。ELM 选择隐层数为 40,迭代次数 20^[17]。结果如图 9 所示,横坐标代表样本数目,纵坐标代表分类器分类的正确率。

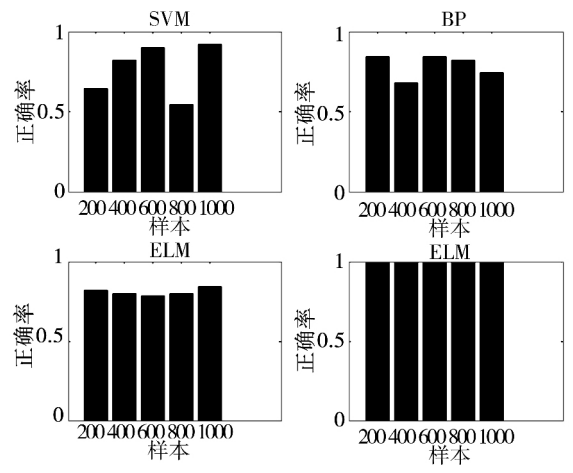


图 9 分类算法正确率分布示意图

由图 9 可以看出,SVM、BP、ELM、DBN 这 4 种分类算法在处理 5 种类别文本时所得到的分类正确率。SVM 分类时性能比较不稳定,5 类样本准确率分布不均匀,BP 和 ELM 分类时性能较稳定,5 类样本的分类正确率比较均匀且数值都较高,但是从图 9 中明显可以看出 DBN 分类性能更优。

图 10 是四种分类算法随着文本数量的递增所显示的分类正确率,横坐标代表文本的测试样本个数,纵坐标代表分类器分类的正确率。本实验五类样本随机各选取 10 个,合计 50 个样本,从图 10 中可看出,每种分类算法是在对第几个样本分类时判断错误,也即分类算法的分类正确率的转折点,对比实验结果可知 4 种分类算法中 DBN 算法分类性能较高。

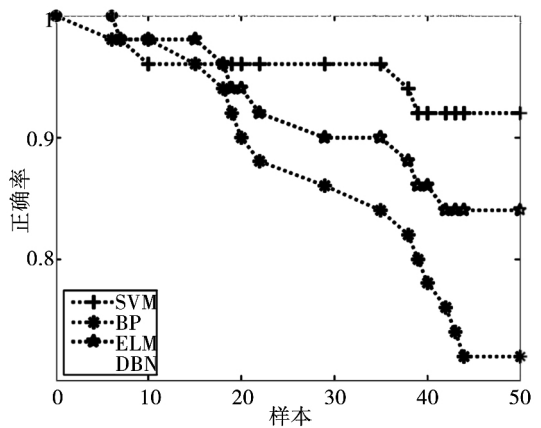


图 10 4 种分类器的分类结果对比示意图

使用 1 000 个测试样本,4 种分类方法正确率对比结果如图 11 所示,该图可以看出每个分类器的分类效果。图 11 横坐标代表测试样本的数目,纵坐标代表分类的正确率。

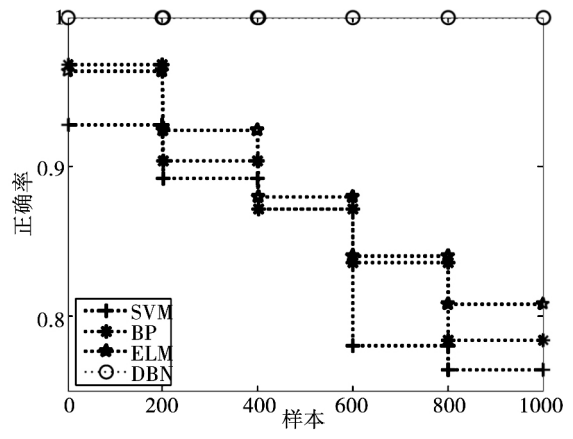


图 11 4 种分类器分类总结果对比示意图

采用相同的训练样本与测试样本,使用 SVM、BP、ELM 这 3 种分类算法与 DBN 分类算法作对比,这 4 种算法实验结果如表 4 所示。

表 4 文本分类算法的结果对比表

		SVM	BP	ELM	DBN
花类	错误数	72	32	36	0
	正确率	64%	84%	82%	100%
树木类	错误数	36	64	40	0
	正确率	82%	68%	80%	100%
虫类	错误数	20	32	44	0
	正确率	90%	84%	78%	100%
土壤类	错误数	92	36	40	0
	正确率	54%	82%	80%	100%
水类	错误数	16	52	32	0
	正确率	92%	74%	84%	100%
综合	正确率	76.4%	78.4%	80.8%	100%

由实验结果可知,本文提出的基于深度学习的文本分类算法能够实现对上面 5 类文本的准确分类,分类效果比基于 BP、SVM、ELM 的分类算法更好,并且基于深度学习的文本分类算法不需要对数据进行降维处理就可自动从高维的原始特征中提取出高度可区分的低维特征,显著提高了分类的效率^[18]。此外,当样本数量达到上万时,基于 BP、SVM、ELM 的分类算法由于样本过多容易产生“过拟合”情况导致分类效果变差,但是 DBN 在面对大数据时表现出了极强的优势,不但可以考虑到文本

更为全面的信息量,而且还可以快速训练分类模型,提高分类效率^[19-20]。

3 结 语

基于深度学习的文本分类算法,文本特征的表达通过 TF-IDF 方法计算权值后得到文本的初始特征矩阵来实现,之后进行 DBN 训练来构建分类器,并通过对其优化来实现文本的快速精准分类。实验结果表明文本的分类可采用深度学习的算法,分类正确率明显高于 BP、SVM、ELM 分类算法,为文本分类提供了新思路^[8]。

参 考 文 献:

- [1] 陈宇,许莉薇. 基于高斯混合模型的林业信息文本分类算法[J]. 中南林业科技大学学报 2014, 34(8):114-119.
- [2] 陈翠平. 基于深度信念网络的文本分类算法[J]. 计算机系统应用 2015, 24(2):121-126.
- [3] 张浩,汪楠. 文本分类技术研究进展[J]. 计算机与信息技术, 2007, 23(1):95-96.
- [4] 柳长源,毕晓君,韦琦. 基于向量机学习算法的多模式分类器的研究及改进[J]. 电机与控制学报 2013, 17(1):114-118.
- [5] 李东洁,李君祥,张越,等. 基于 PSO 改进的 BP 神经网络数据手套手势识别[J]. 电机与控制学报 2014, 18(8):87-93.
- [6] 仲伟峰,马丽霞,何小溪. PCA 和改进 BP 神经网络的大米外观品质识别[J]. 哈尔滨理工大学学报 2015, 20(4):76-81.
- [7] 李军,乃永强. 基于 ELM 的机器人自适应跟踪控制[J]. 电机与控制学报 2015, 19(4):106-116.
- [8] 陈宇,王明月,许莉薇. 基于 DE-ELM 的林业信息文本分类算法[J]. 计算机工程与设计 2015, 36(9):2412-2431.
- [9] 朱少杰. 基于深度学习的文本情感分类研究[D]. 哈尔滨:哈尔滨工业大学 2014.
- [10] 刘树春. 基于支持向量机和深度学习的分类算法研究[D]. 上海:华东师范大学 2015.
- [11] HINTON GE, SALAKHUTDINOV RR. Reducing the Dimensionality of Data with Neural Networks[J]. Science 2006, 313(5786):504-507.
- [12] 戚孝铭. 基于蜂群算法和改进 KNN 的文本分类研究[D]. 上海:上海交通大学 2013.
- [13] 陈宇,许莉薇. 基于优化 LM 模糊神经网络的不均衡林业信息文本分类算法[J]. 中南林业科技大学学报 2015, 35(4):27-59.
- [14] 段江丽. 基于 SVM 的文本分类系统中特征选择与权重计算算法的研究[D]. 太原:太原理工大学 2011.
- [15] 陈勇,黄婷婷,张开碧,等. 结合 Gabor 特征和深度信念网络的人脸姿态分类[J]. 半导体光电 2015, 36(5):815-819.
- [16] HINTON GE, OSINDERO S, TEH Y. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation 2006, 18(7):1527-1554.
- [17] 唐晓明,韩敏. 一种基于极端学习机的半监督学习方法[J]. 大连理工大学学报 2010, 50(5):771-776.
- [18] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning 2009, 2(1):1-127.
- [19] 侯思耕. 基于主题模型和深度置信网络的文本分类方法研究[D]. 昆明:云南大学 2015.
- [20] 鲁铮. 基于 T-RBM 算法的 DBN 分类网络的研究[D]. 长春:吉林大学 2014.

(编辑:温泽宇)