

基于知识规则的构造性优先排序神经网络算法

谷秧波^{1,2}, 武妍^{1,2}, 朱世交^{1,2}, 王守觉^{2,3}

(1. 同济大学 计算机科学与工程系, 上海 200092; 2. 同济大学 半导体与信息技术研究所, 上海 200092;

3. 中国科学院 神经网络实验室, 北京 100083)

(spiritgby@hotmail.com)

摘要:从高维空间样本点覆盖的角度, 讨论了基于知识规则的构造性优先排序神经网络(PONN)算法的原理, 提出了网络构造过程的一般算法以及基于随机取样规则和重心点规则的两个实例算法。实例算法对螺旋线识别和语种识别进行了仿真。实验结果证明了算法的有效性。语种识别实验结果也表明基于重心规则的 PONN 算法在一定条件下优于 SVM。

关键词:优先度排序神经网络; 知识规则; 覆盖

中图分类号: TP183 **文献标志码:** A

Learning algorithm model of constructive PONN based on knowledge rules

GU Yang-bo^{1,2}, WU Yan^{1,2}, ZHU Shi-jiao^{1,2}, WANG Shou-jue^{2,3}

(1. Department of Computer Science and Engineering, Tongji University, Shanghai 20092, China;

2. Institute of Semiconductors and Information Technology, Tongji University, Shanghai 20092, China;

3. Laboratory of Artificial Neural Networks, Chinese Academy of Sciences, Beijing 100083, China)

Abstract: From the perspective of covering points in high-dimensional space, we discussed the theory of constructive PONN based on knowledge rules. PONN's general constructive algorithm and its two special methods based on random sampling and barycenter rules respectively were proposed in this paper. Simulations on spirals recognition and voice language identification were conducted using the special methods. Experimental results illustrate the good performance of PONN instances. Besides, voice language identification results show that PONN method based on barycenter rule outperforms SVM under certain circumstances.

Key words: Priority Ordered Neural Network (PONN); knowledge rule; coverage

0 引言

目前, 使用和研究最多的神经网络是 BP 网络及其改进算法^[1-2], 误差梯度下降算法使前馈神经网络的实际应用前进了一大步, 但是仍有些理论问题未解决^[3], 如难以确定层数和各层节点数, 内部学习过程难以理解, 学习过程仍是一黑箱, 难以控制^[4-5]。近年来, 基于统计理论的 SVM 方法也得到了广泛关注, SVM 通过在投影后的高维空间使用类间最大间距进行类别划分^[6-7], 但其依赖的二次优化方法在高维存在困难且缺乏类似于神经网络的生物学解释。文献[8-9]从人的认识事物的角度提出了仿生模式识别方法, 把模式识别问题抽象成高维空间形体的学习过程, 强调事物认识过程中的联系。

1 优先度排序神经网络的基本原理

1.1 相关符号定义

为描述方便, 现将基本符号定义说明如下。

\mathbf{R}^N : 表示 N 维欧氏空间, \mathbf{Z}_+ 为正整数集。

X_i : 表示 n 维空间中第 i 特征类子空间, $X_i \subset \mathbf{R}^N, i = 1, \dots, m$; 样本点 $x_j^{(i)} \in X_i$ 。

D : 表示当前训练样本集合, $D = \cup X_i$ 。

$\{\psi_i^k\}$, 第 i 类、 k 优先度的神经元集合, $k \in \mathbf{Z}_+, i = 1,$

\dots, m 。单个神经元 $\psi_{i,j}^k$, 表示 i 类 k 优先度的第 j 个神经元。

L_k : 表示构造得到的 k 优先度神经元集合, 定义

$L_k := \{\{\psi_i^k\} \mid k \in \mathbf{Z}_+, i = 1, \dots, m\}$ 。

O_k : 神经元 k 层输出序列, 定义为: $O_k := \{0, \dots, 1_k^p, 0, 0 \mid k = 1, \dots, m, p = 0 \text{ 或 } 1\}$

Y_i : 类别 i 的判别输出, 表示是否激活。

1.2 PONN 网络基本结构

优先度排序网络的出发点是通过高维空间点之间的关系, 使用复杂形体对样本点集进行外覆盖, 每次覆盖神经元具有不同的优先度, 空间神经元覆盖具有不同层次级别, 这样神经元之间不仅通过权值进行知识的表示, 而且包含相互之间的优先关系。通过这种关系可以进行知识的积累和发现, 网络这种认知能力更加符合人类学习的过程(如在线学习)。

网络结构基本单元由各个不同层次优先度隐含层神经网络单元构成, 其基本层次结构如图 1 所示, 通过不同层次的神经元数据分类映射, 形成一个分类判别的塔式结构网络, 样本在高维空间中的样本分类和识别过程依赖于数据本身而进行的, 并非人为设定。网络通过分析样本在空间中的分布特点进行神经网络的动态构造; 构造好的网络对样本识别过程依赖内部不同层次优先度神经元的并行计算激活状态, 进而用以判别被识别样本的分类。

1.3 网络构造基本原理

考虑实际样本的噪声干扰等因素, 因此考虑单个神经元

收稿日期: 2008-01-17; 修回日期: 2008-03-19。 基金项目: 国家自然科学基金资助项目(60475019)。

作者简介: 谷秧波(1979-), 男, 浙江上虞人, 博士研究生, 主要研究方向: 模式识别、流形学习、图像处理; 武妍(1967-), 女, 山西晋中人, 教授, 博士生导师, 主要研究方向: 神经网络、计算智能、模式识别; 朱世交(1980-), 男, 安徽五河人, 博士研究生, 主要研究方向: 模式识别、神经网络、图像处理; 王守觉(1925-), 男, 江苏苏州人, 院士, 博士生导师, 主要研究方向: 仿生模式识别、神经网络。

$\psi_{i,j}^k$ 在高维特征空间中是有限个样本点的外覆盖区域,外覆盖神经元的构造依据样本点在特征空间的分布,因此覆盖形体可能会很复杂(更多自由度,或不同神经元混合)。此外,考虑到分析问题的方便,认为在高维空间中覆盖区域具有各向同性,样本空间为欧式空间,样本点之间的近邻度量采用 L_2 度量空间,通过计算样本特征点 x 与空间参考点 x_0 距离的比较进行判别,如式(1)所示。

$$d(\psi_{i,j}^k) = \sqrt{\sum_{l=1}^n (x_l - x_{l0})^2} \quad (1)$$

当 $d(\psi_{i,j}^k)$ 度量值小于阈值 $T, T \in \mathbf{R}$ (阈值内不包括异类点,由算法动态决定),则认为 x 在神经元外覆盖区域范围之内,即测试点 x 激活神经元 $\psi_{i,j}^k$; 反之,则表示 x 在覆盖区域之外。判别的形式化描述如式(2)所示。

$$F(d) = \begin{cases} 1, & d \leq T \\ 0, & \text{其他} \end{cases} \quad (2)$$

对于构造得到的神经元 $\psi_{i,j}^k, k$ 优先级 i 类别神经元的外覆盖定义为 $Cover(\psi_i^k)$, 同优先级、不同类神经元的外覆盖区域不重叠, $Cover(\psi_i^k) \cap Cover(\psi_j^k) = \emptyset, i \neq j$, 不同优先级神经元的高维空间外覆盖区域有可能出现重叠,此现象类似于人类在做同一事物判别时,使用不同的角度观测(优先级)会得到不同的结论,同时,样本之间不仅有区别,且有很多联系,例如序列关系、相似度等。

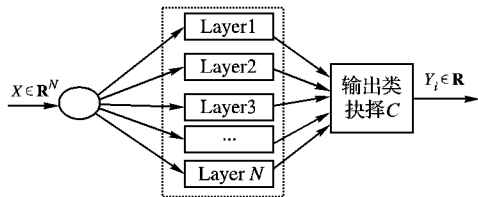


图1 PONN 网络层次构造单元

这里需要说明两点:1) PONN 对网络构造使用的神经元模型没有限制。RBF^[2],超香荔^[7]等封闭式神经元都可以作为覆盖样本空间区域的基本单元使用。在类别覆盖时,本文采用各向同性超球神经元,主要出于模型简单的考虑,且不失一般性。2)为了更好地描述样本在空间的分布,需要分析在不同优先级层次下神经元的连接组合问题,实现神经元对样本空间的有序(优先级)局部区域外覆盖,强化覆盖的精确性。非球神经元以及混合神经元模型的讨论另文描述。

PONN 网络构造的思路。网络构造及训练过程依据训练样本点在高维空间的分布特征。起始阶段 PONN 内部优先级(层次)神经元为空,即 $k = 0$ (不同于传统网络,需预先确定神经元数目);随着训练样本进入网络学习,层 k 逐渐增大,神经元数量不断增多,构成隐含层内神经元 $\{\psi_i^k\}$ 集合,实现对训练样本集合中各类别的覆盖,同时空间外覆盖具有一定泛化能力。

PONN 网络识别的基本方法。对于待识别的数据点 $x \in \mathbf{R}^N$, 网络不同优先级神经元对应响应输出 O_k , 一般考虑优先级高(最新学习的内容)的先被判别,定义判别 $Y_i := \{i | i = \{0 \text{ 或 } \argmin_{k \in \mathbf{Z}_+} O_k > 1\}\}$, 通过 Y_i 对识别样本 $x \in \mathbf{R}^N$ 进行类别判断:

$$Y_i = \begin{cases} \text{类别 } i, & i > 0 \\ \text{拒绝识别}, & i = 0 \end{cases} \quad (3)$$

由于采用空间外覆盖方法进行类别判定,对于未知类别的样本,PONN 并不会出现乱识别情况;而传统网络则把样本空间划分完,不“认识”的样本容易出现乱识别的问题。

1.4 网络构造算法的一般过程描述

定义 1 知识规则对优先度的映射。网络对分类存在知识规则 Rule,通过知识规则映射高维空间 PONN 神经元的优先关系(序关系^{[11]1-3})。

定义 2 优先级序列描述。对于 PONN 网络要获取的知识类别 K , 定义优先级序列 (L_i, K) 作为对分类中知识获取的描述,优先级依赖于样本知识获取规则,针对不同规则同一样本会得到不同的优先级(序关系^{[11]1-3})。

通用学习规则形式化描述如下:

General Algorithm of PONN

输入:样本 $D = \{X_1, \dots, X_m\}$, 优先级 $k = 1$

输出:具有优先级关系的神经网络

While $D \neq \emptyset$ do

For 类别样本 in D

→优先级规则映射 Rule 以及序关系描述

→使用样本进行同类空间区域外覆盖

→把覆盖后的类别样本赋予新 D' 队列

End of for

→ D' 赋予 D , 进行下一优先级覆盖

$K = K + 1$

End while

对优先级规则可能会包括一些复杂的逻辑规则,同时高维空间点的运算规则也有待研究,但此时存在点之间的基本关系,例如近邻关系、连接密度等规则,这些规则是从事物特定应用角度来看待的,而非人为预先规定,本文暂且从这些简单的关系出发进行研究。

1.5 网络构造过程中的收敛性分析

假设 N 和 M 分别为输入和输出的向量维数,分别对应输入向量的维数和输出类别,网络收敛的构造性说明如下,构造性证明不仅能说明神经网络的构造是可能的,最重要的是具有可操作性。

定理 假设当前神经元的优先级为 p , 及其网络整体拟合误差为 e_p , 可以通过添加(或修改)新的较高优先级神经元层 $p + 1$, 使得整体类别拟合误差 $e_{p+1} < e_p$ 。

证明 考虑已构造好的 PONN 网络,假设其他分类都正确,但存在 A 类被误映射为 B 类的情况 $f: Cover_p(\langle x_1, x_2, \dots, x_n \rangle_A) \rightarrow B$, 假设误差为 e_p , 此空间区域被空间神经元 B 类外覆盖所覆盖,通过添加新的神经元覆盖 $f: Cover_{p+1}(\langle x_1, x_2, \dots, x_n \rangle_A) \rightarrow A$ (几何描述如图 2 所示),使得误识区域被新覆盖所代替,因为 $(p + 1) > p$, 当网络进行识别时,错误的样本被新神经元覆盖住,从而保证样本被正确识别,其误差为 e_{p+1} , 当前误差定有 $e_{p+1} < e_p$ 。考虑另外一种情况就是调整空间神经元的外覆盖区域,把 $Cover_p$ 替换为 $Cover'_p$, 并满足在原有样本覆盖正确的条件下使得 $f: Cover'_p(\langle x_1, x_2, \dots, x_n \rangle_A) \rightarrow A$, 此过程可以通过算法的有限次迭代完成,得到有限样本覆盖区域,并且算法收敛。

2 神经网络学习的构造算法

针对 PONN 网络原理,使用点之间的规则,给出两个网络构造算法的实现,方法侧重于网络训练过程,使用超球类型的神经元。

2.1 随机分割算法(RPA)

随机分割算法(Random Partitioning Algorithm, RPA)分别随机选取训练样本集中不同类别的样本点作为支撑点进行空间近邻点的覆盖,算法描述如下:

- 1) 初始化 $i = 1, i \in \{1, 2, \dots, m\}, k = 1$ 。
- 2) 从当前层次的样本队列中, 随机选取测试样本点 $x \in X_i, j = 0$ 。
- 3) 以 $x \in X_i$ 为支撑点, 使用球形 $\psi_{i,j}^k$ 神经元外覆盖 X_i , 覆盖得到 X_i 的子集合为 $X_{sub}, X_{sub} \subseteq X_i \circ X_i = X_i \setminus X_{sub}$, 如果 $X_i = \emptyset$, 则从样本队列中去除 X_i 类别; 否则, X_i 继续进入下一层样本队列。
- 4) 判别当前优先度样本队列是否为空, 如果为空, 则执行第5步; 否则 $i = i + 1$, 转第2步。
- 5) 判别下一层次样本队列是否为空, 若是, 则构造结束; 否则 $k = k + 1$, 转第2步。

RPA 算法样本点的选择没有利用任何先验知识, 以随机选取的样本点作为支撑点进行高维空间外覆盖。由于支持点是随机的, 算法可能是不稳定的。

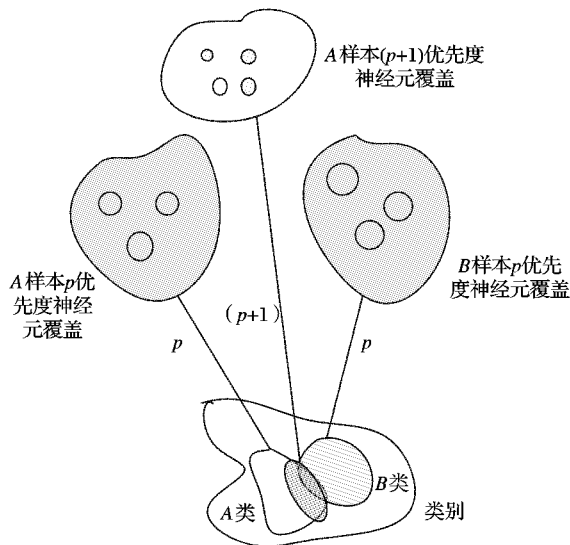


图2 (P+1) 优先度神经元空间覆盖的几何解释

2.2 中心适配选择算法 (CASA)

中心适配选择算法 (Center Adaptive Selection Algorithm, CASA), 基本思路就是利用各个样本点相应的中心聚集点, 一种方法是取同类样本的重心点, 但重心点很可能不是真实样本点, 在高维空间点稀疏的情况下更是如此, 因此可通过调整重心为实际样本点, 使调整后的点作为支撑点, 进行空间区域外覆盖的构造。算法描述如下:

- 1) 初始化 $i = 1, i \in \{1, 2, \dots, m\}, k = 1$ 。
- 2) 从当前层次的样本队列中, 选取测试样本空间重心点 $x_0 = \frac{1}{|X_i|} \sum_{i=0}^{|X_i|} x_i, x_i \in X_i, j = 0$ 。
- 3) 以 x_0 为支撑点, 使用 $\psi_{i,j}^k$ 球形神经元覆盖 X_i , 覆盖得到的 X_i 的子集合为 $X_{sub}, X_{sub} \subseteq X_i$ 。
- 4) 若 $X_{sub} = \emptyset$, 则选取 $x_0 = \{x \mid \max(d(x_0, x)), x \in X_i\}$, 转第3步。
- 5) 选取 $X_i = X_i \setminus X_{sub}$, 如果 $X_i = \emptyset$, 从样本队列中去除 X_i 类别; 否则, X_i 继续进入下一优先度样本队列。
- 6) 判别当前层次样本队列是否为空, 如果为空, 则执行第7步; 否则 $i = i + 1$, 转第2步。
- 7) 判别下一层次的样本队列是否为空, 若是, 则结束; 否则 $k = k + 1$, 转第2步。

CASA 作为网络学习的特定实现, 利用样本重心点规则寻找各个类别的支撑点, 算法仍有改进余地。

3 试验结果与讨论分析

实验包括二维螺旋线和语种识别算法的比较。

3.1 螺旋线

三条螺旋线定义如式(4), 平面上图形如图3所示。

$$\begin{cases} R_1 = 0.4 \times \theta, \theta \in [-0.4\pi, 4\pi] \\ R_2 = 0.4 \times (\theta + \pi), \theta \in [-0.5\pi, 5\pi] \\ R_3 = 0.4 \times (\theta + 1.7 \times \pi), \theta \in [-0.6\pi, 6\pi] \end{cases} \quad (4)$$

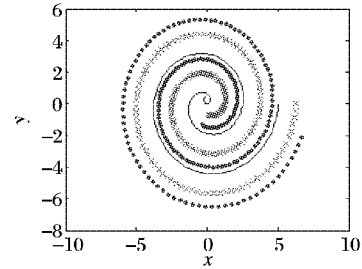


图3 平面中三条螺旋线图像

为了观测不同训练样本对网络构造的影响, 训练时每类分别随机取 150 ~ 350 个数量范围, 间隔为 5, 测试每类取 1000 个样本点。

图4描述了不同训练样本数量下 RPA 和 CASA 对于螺旋线的神经网络构造情况, CASA 构造的神经元比 RPA 少。图5则是两种算法识别情况的比较。从中可看出: CASA 算法随着训练样本的不同, 误识率上升变化也较明显, 因为 RPA 使用的是随机样本选择, 因此易表现出不稳定性。

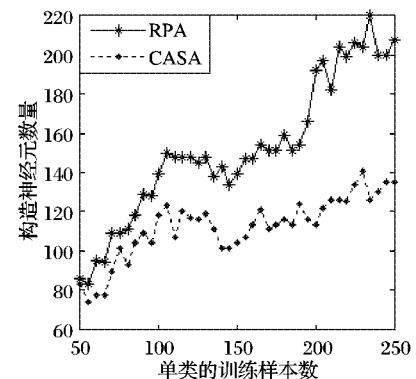


图4 RPA 和 CASA 构造神经元数量变化

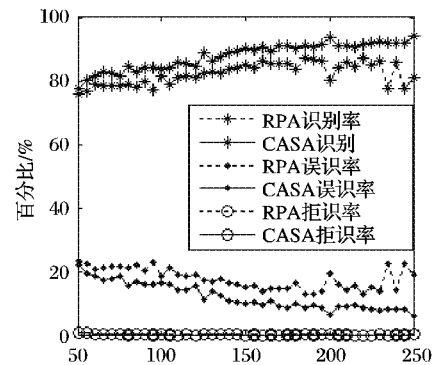


图5 RPA 和 CASA 对螺旋线的识别变化

3.2 语种识别

对不同时间片长语音的语种识别, 采用语音的 MFCC^[10] 特征进行训练和测试。采集到的不同片长的各类语音情况如表1所示。表2是 CASA 和 SVM 在不同片长情况下的语种语音识别率比较, 表2中的维数指的是表1不同片长语音的 MFCC 特征维数。SVM 的核函数采用 RBF 核, 具体的参数串

为“-s 0 -c 1000 -t 2 -wi 1”,相关参数的意义和取值参见 LIBSVM^[6]对应的网站,其他参数取默认值。

表 1 语种语音样本的分布情况

样本	时间/s	样本数量			
		中文	英文	日文	总计
片长 1	1.008	2323	4254	1007	7584
片长 2	1.504	1325	2449	623	4397
片长 3	2.000	827	1611	427	2865
片长 4	2.496	567	1119	315	2001

表 2 语种语音样本的识别比较

维数	PONN			
	中文	英文	日文	平均
992	91.031	84.177	95.327	87.562
1488	91.928	84.177	88.785	86.940
1984	89.238	76.160	80.374	80.348
2480	77.130	86.709	83.178	83.831
维数	SVM			
	中文	英文	日文	平均
992	89.256	84.758	93.443	87.140
1488	92.562	82.156	95.082	86.696
1984	90.083	75.836	86.885	81.153
2480	81.818	90.706	80.328	87.140

通过比较 PONN 和 LIBSVM 的语种语音识别结果发现, CASA 算法在一定情况下(如表 2 中带下划线数据所示)要比 SVM 好,具有一定竞争力,但不是绝对。但由于 SVM 方法转化为最优化问题的求解,在高维数据情况下难以计算,而 PONN 方法利用同类、异类样本点集关系,更具可计算性,具体时间复杂度的分析有待进一步研究。

4 结语

本文研究了 PONN 网络的学习算法框架,提出一般学习过程描述,对特定算法 RPA、CASA 进行了描述、分析,通过实验可以看出:

1)结合事物高维空间分布特点,有利于提高网络构造性能力和泛化能力。

2)简单的 CASA 测试结果在一定情况下比 SVM 好,具有很强的竞争力。

PONN 网络通过动态构造方法进行学习,把网络结构与神经元分开,从而克服传统网络神经元数量难以确定、收敛速度慢等问题。如何更好地分析高维空间点之间的关系,构造合适的网络结构,是进一步要解决的问题。

参考文献:

- [1] HAYKIN S. 神经网络原理[M]. 北京:机械工业出版社,2004.
- [2] BAKRY H M, MASTORAKIS N. New fast normalized neural networks for pattern detection[J]. Image and Vision Computing, 2007, 25(11): 1767-1784.
- [3] ISLAM M M, YAO X, MURASE K. A constructive algorithm for training cooperative neural network ensembles[J]. IEEE Transactions on Neural Network, 2003, 14(4): 820-834.
- [4] LIU T, Moore A W, Gray A. New algorithms for efficient high-dimensional nonparametric classification[J]. Journal of Machine Learning Research, 2006, 7(6): 1135-1158.
- [5] PAREKH R, YANG J H, HONAVAR V. Constructive neural network learning algorithm for pattern classification[J]. IEEE Transactions on Neural Network, 2000, 11(2): 436-451.
- [6] CHANG C C, LIN C J. LIBSVM-A library for support vector machines [EB/OL]. [2007-10-10]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [7] IKEDA K, YAMASAKI T. Incremental support vector machines and their geometrical analyses[J]. Neurocomputing, 2007, 70(13/15): 2528-2533.
- [8] WANG S J. Blomimetic pattern recognition(in Chinese)[J]. Acta Electronica Sinica, 2002, 30(10): 1417-1420.
- [9] WANG SHOU-JUE, LAI JIANG-LIANG. Geometrical learning, descriptive geometry, and biomimetic pattern recognition[J]. Neurocomputing, 2005, 67: 9-28.
- [10] KIM S, ERIKSSON T, KANG H G, et al. A pitch synchronous feature extraction method for speaker recognition[C]// Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Montreal, Canada, 2004, 1: 405-408.
- [11] INGER I M, THORPE J A. 拓扑学与几何学基础讲义[M]. 干丹岩,译. 上海:上海科学技术出版社,1985:1-3.

(上接第 1644 页)

调整的规则,而且使用增量学习方法,保留支持向量,抛弃非支持向量,有效地克服了传统 TSVM 算法中训练速度慢,训练准确率低等问题,具有现实的推广意义。实验结果表明, IPTSVM 算法在各种样本分布情况下都取得了较好的分类效果。

参考文献:

- [1] BURGESS C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [2] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 2000.
- [3] 陈毅松,汪国平,董士海. 基于支持向量机的渐进直推式分类学习算法[J]. 软件学报, 2003, 14(3): 451-460.
- [4] VAPNIK V N. Statistical learning theory[M]. New York: John Wiley and Sons, 1998.
- [5] WANG Y E, HUANG SHANG-TENG. Training TSVM with the proper number of positive samples[J]. Pattern Recognition Letters, 2005, 26(14): 2187-2194.

- [6] JOACHIMS T. Transductive learning via spectral graph partitioning [C]// Proceeding of The Twentieth International Conference on Machine Learning. Washington, DC: [s. n]. 2003: 290-297.
- [7] ZELIKOVITZ S. Transductive LSI for short text classification problems[C]// Proceedings of 17th International Florida Artificial Intelligence Research Symposium Conference. Miami: AAAI Press, 2004.
- [8] JOACHIMS T. Transductive inference for text classification using support vector machines[C]// Proceedings of the 16th International Conference on Machine Learning (ICML). San Francisco: Morgan Kaufmann Publishers, 1999: 200-209.
- [9] CHEN YI-SONG, WANG GUO-PING, DONG SHI-HAI. Learning with progressive transductive support vector machine[C]// Second IEEE International Conference on Data Mining(ICDM'). Maebashi Japan: IEEE Press, 2002: 67-74.
- [10] RATSABY J. Incremental learning with sample queries[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 883-888.