

文章编号: 1003-0077(2010)05-0056-06

基于 Stacking 组合分类方法的中文情感分类研究

李寿山, 黄居仁

(香港理工大学 中文及双语学系, 中国 香港)

摘要: 情感文本分类(简称情感分类)是一种面向主观信息分类的文本分类任务。目前, 由于其广泛的应用前景, 该任务在自然语言处理研究领域得到了普遍关注, 相继出现多种用于情感文本分类的有监督的分类方法。该文具体研究四种不同的分类方法在中文情感分类上的应用, 并且采用一种基于 Stacking 的组合分类方法, 用以组合不同的分类方法。实验结果表明, 该组合方法在所有领域都能够获得比最好基分类方法更好的分类效果。从而克服了分类方法领域依赖的困境(不同领域需要选择不同基分类方法才能获得更好的分类结果)。

关键词: 计算机应用; 中文信息处理; 情感分类; 组合分类器

中图分类号: TP391

文献标识码: A

Chinese Sentiment Classification Based on Stacking Combination Method

LI Shoushan, HUANG ChuRen

(The Hong Kong Polytechnic University, Department of Chinese & Bilingual Studies, Hong Kong, China)

Abstract Sentiment-based text categorization (for short, sentiment classification) is a task of classifying text according to the subjective information in the text. Nowadays, it has been closely studied in the research field of natural language processing (NLP) due to its wide real applications. As a result, many supervised machine learning classification approaches have been applied to this task. In this paper, we research on four classification approaches and propose a new combination method based on stacking to combine these four approaches. Experimental results show that our combination method achieves better performances than the best single one. Therefore, this combination method can avoid selecting a suitable classification approach according to different domains.

Key words: computer application; natural language processing; sentiment classification; multiple classifier combination

1 引言

目前, 人们越来越习惯于在网络上表达自己的观点和情感。网络上出现了大量的带有情感信息的文本, 这些情感文本以商品评论、论坛评论、博客等形式存在。面对这些越来越多表达情感信息的文本, 传统的基于主题的文本分类系统已经不能满足人们的需求, 迫切需要对这些情感文本进行研究和分析。情感文本分类即是在这种氛围下产生的一种特殊的文本分类任务, 该任务按照表达的情感倾向性对文本进行分类^[1]。例如, 判断文本对某个事物

的评论是“好”还是“坏”。该任务的研究历史虽然不长, 但是已经成为自然语言处理方向里面的一个研究热点。尤其是近几年在自然语言处理相关国际顶级会议上(ACL、EMNLP、SIG IR、WWW 等)涌现出了大量的文章。目前, 该方向的研究主要不仅仅是在英文方面, 中文情感文本分类的研究也得到了迅速的发展, 例如, 文献[2-4]。

针对情感分类任务, 目前主流的方法是基于机器学习的分类方法, 该方法利用统计机器学习分类方法学习标注样本, 然后用学习好的分类器测试非标注样本。这种方法在性能上比其他的基于规则的方法有着明显的优势^[1]。不过, 机器学习方法中存

收稿日期: 2009-08-27 定稿日期: 2009-12-24

作者简介: 李寿山(1980—), 男, 博士后, 研究方向为自然语言处理、模式识别; 黄居仁(1958—), 男, 讲座教授, 研究方向为自然语言处理、语言学。

在多种不同的分类方法,选择合适的分类方法便成为情感分类研究中的一个重点问题。文献[4]首先采用机器学习方法进行英文文本(电影评论)的情感分类研究,文中给出三种不同的分类方法,分别为朴素贝叶斯(Naive Bayes, NB),最大熵(Maximum Entropy, ME)和支持向量机(Support Vector Machines, SVM)。他们的研究结果发现大部分的情况下 SVM 取得的分类效果最好。后续的许多研究工作都是基于 SVM 的分类方法进行展开的,例如,文献[5-6]。对于中文文本的情感分类研究,文献[8]采用 K-近邻(KNN),感知器 Winnow, NB 和 SVM 四种分类方法对来自不同领域的评论文本进行分类,实验结果发现 SVM 在所有领域都取得了最好的分类效果。文献[9]也有类似的结论,不过实验结果中 SVM 的分类效果没有非常大的优势,而且在利用不同的特征时, SVM 的效果有时候会差于其他分类方法。然而,文献[10]给出了的结论与上面的文章不是很一致,该文发现如果利用二元(Bigram)词特征或者单字特征的时候, NB 的分类效果是最好的。因此,我们可以认为 SVM 分类是情感文本分类任务中表现比较好的分类方法,但是该方法不可能在所有领域所有特征集合下取得最好的分类效果。

本文将采用四种不同的分类方法,即朴素贝叶斯,最大熵,支持向量机和随机梯度下降线性分类方法对中文情感文本分类进行研究。本文的重点不是比较这四种方法的分类表现。我们相信不同的领域或者不同的特征可能需要不一样的分类方法才能取得最好的分类效果。因此,我们的目标是应用组合分类方法组合这四种方法,考察组合后的方法能否获得比最好分类方法更好的分类效果。

本文的其余部分如下安排:第2节详细介绍各种分类方法在情感分类上面的应用;第3节给出基于 Stacking 的组合分类器方法;第4节是实验结果和分析;最后一节是本文的结论和将来的一些工作。

2 统计情感文本分类方法

2.1 朴素贝叶斯(Naive Bayes, NB)

该方法的基本思想是利用特征项和分类的联合概率来估计给定文档的分类概率^[11]。朴素贝叶斯假设文本是基于词的 Unigram 模型,即文本中词的出现依赖于文本类别,但不依赖于其他词及文本的

长度,也就是说,词与词之间是独立的。文档一般采用 DF 向量表示法,即文档向量的分量为一个布尔值:0 表示相应的单词在该文档中未出现,1 表示出现。则文档 d 属于类 c 文档的概率为:

$$P(c|d) = \frac{P(c) \prod_{t \in V} P(d(t)|c)}{\sum_c P(c) \prod_{t \in V} P(d(t)|c)} \quad (1)$$

其中,

$$P(d(t)|c) = \frac{1 + N(d(t)|c)}{2 + |d_c|} \quad (2)$$

其中, $P(d(t)|c)$ 是对在 c 类文档中特征 t 出现的条件概率的拉普拉斯估计, $N(d(t)|c)$ 是 c 类文档中特征 t 出现的文档数, $|d_c|$ 为 c 类文档所包含的文档的数目。对于情感文本分类, $c \in \{-1, 1\}$, 分别表示负面(贬)或者正面(褒)。

2.2 最大熵(Maximum Entropy, ME)

最大熵分类方法是基于最大熵信息理论,其基本思想是为所有已知的因素建立模型,而把所有未知的因素排除在外。也就是说,要找到一种概率分布,满足所有已知的事实,但是让未知的因素最随机化。相对于朴素贝叶斯方法,该方法最大的特点就是不满足特征与特征之间的条件独立。因此,该方法适合融合各种不一样的特征,而无需考虑它们之间的影响。

在最大熵模型下,预测条件概率 $P(c|d)$ 的公式如下:

$$P(c|d) = \frac{1}{Z(d)} \exp \left(\sum_i \lambda_{ic} F_{ic}(d, c) \right) \quad (3)$$

其中, $Z(d)$ 是归一化因子。 F_{ic} 是特征函数,定义为:

$$F_{ic}(d, c') = \begin{cases} 1, & n_i(d) > 0 \text{ and } c' = c \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

在做情感分类的时候,我们用的特征主要是词特征是否出现在某个文档里并属于某个类别。

2.3 支持向量机(Support Vector Machines SVM)

该方法主要是用于解决二元分类模式识别问题,它的基本思想是在向量空间中找到一个决策平面(Decision surface),这个平面能“最好”地分割两个分类中的数据点^[12]。支持向量机的核心思想就是要在训练集中找到具有最大类间界限的决策平面。

支持向量机方法中一个有趣的特性是决策平面

只是那些刚好和决策平面距离为 $1/\|\vec{w}\|$ 的数据点来决定, 称为支持向量, 删除其他数据点是不会影响算法结果的(即产生的决策函数不变), 这点是它与其他分类方法的不同之处之一, 令它可以一定程度上避免“维数灾难”。因此, 该方法比较适合那些高维向量空间的分类问题。这一点正好满足情感文本分类的需求。长期以来, 该分类方法在情感文本分类任务中扮演着非常重要的角色^[5-7]。

2.4 随机梯度下降 (Stochastic Gradient Descent, SGD)

随机梯度下降是一种线性分类方法^[13], 即 $f(d) = w^T d$ 。其中 d 是文档的向量表示。该分类方法的训练目标就是求得最佳的参数 w 如下:

$$\hat{w} = \arg \min_w \sum_{x,y} L(w^T d, c) + \frac{\lambda}{2} \|w\|_2^2 \quad (5)$$

其中, $\|w\|_2^2 = w^T w$, λ 是一个非负的正规化因子。如果 $\lambda = 0$, 表明该问题是非正规问题。

顾名思义, 随机梯度下降分类方法是利用随机梯度下降解决公式(5)中表达的优化问题, 从而得到 SGD 的在线更新策略如下^[13]:

$$\hat{w}_t = \hat{w}_{t-1} - \eta S^{-1} (\lambda \hat{w}_{t-1} + L'_1(\hat{w}_{t-1}^T d_t, c_t), d_t) \quad (6)$$

其中 $L'_1(p, y) = \frac{\partial}{\partial p} L(p, y)$, (d_t, c_t) 表示样本

在第 t 步的分类结果 c_t 。矩阵 S 可以认为是一个先决调节器 (Pre-conditioner)。简单起见, 该矩阵设置为一个常数矩阵。 $\eta > 0$ 是学习速率调节参数。该分类算法的流程如下^[13]:

```

初始化  $\hat{w}_0$ 
for  $t = 1, 2, \dots$ 
    从样本集合里面随机抽取样本  $(d_t, c_t)$ ;
    更新  $\hat{w}_{t-1}$  如下:
         $\hat{w}_t = \hat{w}_{t-1} - \eta S^{-1} (\lambda \hat{w}_{t-1} + L'_1(\hat{w}_{t-1}^T d_t, c_t), d_t)$ ;
end
```

该分类方法适用于高维大规模的分类问题^[13]。文献[14]将该算法应用于英文的情感文本分类任务中, 取得了同 SVM 类似的分类效果。

3 组合分类方法

组合分类方法是融合多个分类器的结果从而得到一个新的融合结果, 这个融合结果将取代各个基分类器的结果作为最终的分类决定^[15]。组合分类

器方法是模式识别以及机器学习理论研究领域里面的一个重要的研究方向。

3.1 产生基分类器

构建一个多分类器系统, 首先需要产生多个基分类器用以组合。产生基分类器的方式大致有三种。第一种方式是通过训练不同的语料库产生不同的分类器; 第二种方式是通过训练不同的特征集产生多个基于不同特征集的分类器。例如: 在生物认证问题中, 作为认证的特征可以是人脸、声音、虹膜等。这些特征中的任一特征都可以训练单个分类器, 这样, 一个生物认证系统可能包括多种生物特征的分类器。第三种方式是通过不同的学习方法获得不一样的分类器。很多分类方法是基于不同的机理的, 如 K-近邻 (KNN) 方法是基于记忆的方法, 支持向量机方法 (SVM) 是基于结构风险最小理论的方法等等。因此, 不同的分类方法实现的分类器实现分类的效果往往是不一样的。一种可行的方式就是将多个分类方法实现的分类器组合起来。

3.2 融合算法

获得了基分类器的分类结果后, 组合分类器方法需要特别的融合方法去融合这些结果。融合方法本身就是“多分类系统”研究领域的一个基本问题之一^[15]。融合方法可以分为两种, 固定的融合方法 (Fixed Rules) 和可训练的融合方法 (Trained Rules)。前者的优势在于它们不需要额外的训练语料进行训练。这种方法简单, 易实现, 如简单的投票规则 (Voting Rule) 和乘法规则^[15]。可训练的融合方法的优势在于在足够的训练语料下, 它们能够获得更好的分类效果。

假设有 R 个参加组合的分类器 $f_k (k = 1, \dots, R)$, 这些分类器给样本 x 的分类结果为 $L_k (L_k = c_1, \dots, c_m)$ 。另外, 他们提供出了属于每个类别的概率信息: $P_k = [p(c_1 | d_k), \dots, p(c_m | d_k)]^T$, 其中 $p(c_i | d_k)$ 表示样本 d_k 属于类别 c_i 的概率。如果样本 d_k 属于类别 c_j , 在不同的融合算法中需要满足不一样的条件。下面是一种常见的固定融合算法:

乘法规则的条件^[15]:

$$\text{assign } y \rightarrow c_j$$

$$\text{乘法规则: } j = \arg \max_i p(c_i) \prod_{k=1}^R p(c_i | d_k) \quad (7)$$

目前比较主流的一种可训练的融合方法叫做元学习的方法 (Meta-learning)。元学习的融合方法是

指, 将基分类器输出的分类结果作为中间特征^[16] (又叫元特征), 即:

$$P_{meta} = [P_1, P_2, \dots, P_k, \dots, P_m] \quad (8)$$

然后把这些特征向量作为输入再一次学习一个分类器, 该第二层的分类器叫元分类器。

3.3 基于 Stacking 的组合分类器方法

基于 Stacking 的组合分类器方法是目前比较主流的组合分类器方法^[17]。该方法产生基分类器的方式即是上面提到的第三种方式。利用上面提到的四种分类方法, 即 NB, ME, SVM 和 SGD, 训练得到四个基分类器。然后, 使用的融合算法是可训练的元学习方法。特殊的是中间样本(元学习样本)是通过训练样本的 N-fold 交叉验证 (Cross validation) 获得的。另外, 针对元分类器的分类方法也可以有很多种选择, 在后面的实验中, 我们分别选取 SVM 和 ME 做相关实验。基于 Stacking 的组合分类器方法的系统框架结构如图 1 所示。

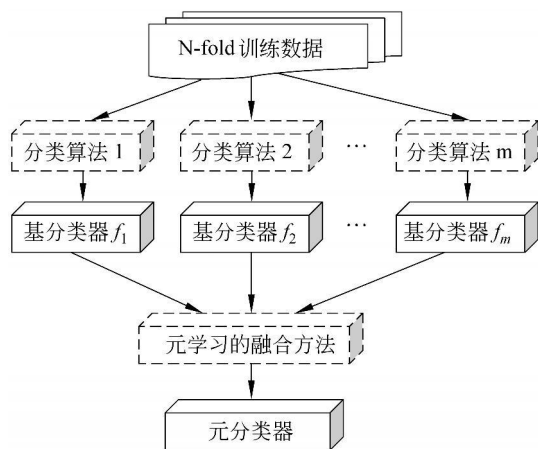


图1 基于 stacking 的组合分类器系统的框架结构

具体来说, 假设 x' 表示开发集中某一个样本的文本向量, 其中开发集是指 N-fold 交叉验证过程中被当作测试集(包括 1fold 样本)的那部分样本。第 l 个基分类器(由其他 $N-1$ 个 fold 的样本训练得到) f_l 在该样本上面的分类结果 $\vec{P}(x'_l)$ 为属于两个类别 $\{c_1, c_2\}$ 的后验概率, 即,

$$\vec{P}(x'_l) = \langle p(c_1 | x'_l), p(c_2 | x'_l) \rangle \quad (9)$$

所有基分类器的分类结果组合构成元分类器的输入向量 $x^{meta} \in R^{2 \times l}$

$$x^{meta} = \langle \vec{P}(x'_1), \vec{P}(x'_2), \dots, \vec{P}(x'_l) \rangle \quad (10)$$

交叉验证使得获得的元学习向量数目同整个训练样本里面的样本数目是一样的。由于元分类器的向量维度非常小(为 $2 \times l$), 该分类器的训练和测试

速度非常快。因此, 元学习过程对系统的时间复杂度影响不大。

4 实验设计及分析

本实验中, 我们使用了谭松波博士收集的一个的中文情感文本分类语料库^①, 该语料库的来源是旅馆预订领域, 我们选取了 2 000 个样本。为了确保实验分析更加可靠, 我们在卓越网站^②上面另外收集了来自三个领域的中文评论语料。这三个领域分别是书籍、DVD 和运动产品。实验中的情感文本分类的任务是将评论分为正面和负面, 每个领域大概有 1 600 个样本。实验过程中, 我们选择 80% 的样本作为训练样本, 剩余的 20% 样本作为测试样本。上面提到的四种分类方法中, SVM 是使用标准工具 light-SVM^③, NB 和 ME 使用的是 MALLET 机器学习工具包^④。在使用这些工具的时候, 所有的参数都设置为它们的默认值, 例如 SVM 使用的是线性核函数。SGD 方法是由我们自主实现的(参考文献[13])。在实现过程中, 我们采用的学习速率为 $\eta = 0.002$, 损失函数 $L(x, y)$ 为 Huber 损失函数^[18]。在实现基于 Stacking 的组合分类方法时, 我们在训练样本中进行了 5-fold 交叉验证来得到元学习样本。然后分别使用 SVM 和 ME 作为元分类器的分类算法。为了便于比较, 我们同样给出了一种固定方式的融合算法—乘法规则的融合结果。

我们采取了分类正确率来评价分类的效果, 其定义如下:

$$A_c = \frac{\text{number of correctly classified samples}}{\text{total number of all samples}} \quad (11)$$

在进行分类之前, 首先我们采用中国科学院计算技术研究所的分词软件 ICTCLAS 对中文文本进行分词操作。给定分好词的文本后, 我们分别选取词的 Unigram 和 Unigram+Bigram 作为特征, 用以获得文本向量的表示。

表 1 和表 2 给出了四个基于不同分类算法的基分类器在四个领域里面的分类结果。表 1 和表 2 的结果分别是利用词的 Unigram 和 Unigram+Bigram

① <http://www.searchforum.org.cn/tansongbo/>

② <http://www.amazon.cn/>

③ <http://svmlight.joachims.org/>

④ <http://mallet.cs.umass.edu/>

表 1 基分类器在各个领域上面的分类结果
(利用词的 Unigram 特征)

领域	SGD	ME	NB	SVM
书籍	0.769	0.785	0.785	0.762
DVD	0.752	0.759	0.740	0.765
运动	0.858	0.849	0.858	0.861
旅馆	0.860	0.850	0.853	0.855

表 2 基分类器在各个领域上面的分类结果
(利用词的 Unigram+Bigram 特征)

领域	SGD	ME	NB	SVM
书籍	0.775	0.794	0.801	0.798
DVD	0.762	0.768	0.774	0.737
运动	0.870	0.864	0.876	0.867
旅馆	0.875	0.882	0.891	0.878

作为特征得到的。从表 1 的结果可以看出,在使用 Unigram 的时候,没有一种分类算法是有绝对优势的。SVM 的表现相对来说比较好,它在两个领域中取得了最好的分类效果。然而,从表 2 的结果可以看出,NB 分类方法取得了一致最好的效果。这一点同文献[10]的结论一致。比较表 1 和表 2 的结果,我们可以发现,引入词的 Bigram 对分类的性能提高有明显的帮助。总体来说,该实验结果同我们

表 3 不同融合算法在各个领域里面的表现
(利用词的 Unigram 特征)

领域	基分类器 最好结果	乘法规则	Stacking (SVM)	Stacking (ME)
书籍	0.785	0.788	0.811	0.808
DVD	0.765	0.778	0.781	0.778
运动	0.861	0.870	0.868	0.877
旅馆	0.860	0.880	0.873	0.878

表 4 不同融合算法的识别结果比较
(利用词的 Unigram+Bigram 特征)

领域	基分类器 最好结果	乘法规则	Stacking (SVM)	Stacking (ME)
书籍	0.801	0.805	0.798	0.811
DVD	0.774	0.768	0.784	0.787
运动	0.876	0.874	0.880	0.886
旅馆	0.891	0.894	0.900	0.900

的观点一致,就是说不同的领域,不同的特征集合需要的最优分类算法往往也不一样。

表 3 和表 4 给出了组合分类方法在四个领域里面的分类结果。表 1 和表 2 的结果分别是利用词的 Unigram 和 Unigram+Bigram 作为特征得到的。为了便于比较 Stacking 组合方法和其他方法,我们还给出了基分类器中获得的最好分类结果和乘法规则融合算法的结果。从表 3 和表 4 的结果可以看出,乘法融合算法基本能够保证组合后的结果接近或者超过最好基分类器的结果(在 DVD 领域使用 Unigram+Bigram 特征时有 0.6% 的性能损失)。然而,基于 Stacking 的组合方法能够很好地提高分类效果,比最好的基分类器的性能都有所提高。这一点对实际应用非常重要,因为我们可以不必去面对每个领域选择合适的分类算法,而是利用组合的方法去组合不同的分类方法。这样组合的结果还可以超过最好的基分类器的结果。一般情况下,ME 分类方法作为元学习的时候,组合分类的结果表现比较稳定。

组合分类方法比最佳分类方法结果平均提高了一个百分点左右,虽然这样的提高并不算非常显著,但是在实际应用中,几乎不可能有一种分类方法能够在不同的领域或者不同的特征集上都能取得最佳的效果(例如表 1 中显示,ME 方法在书籍领域取得最佳效果而在旅馆领域,SGD 方法表现最好)。更重要的是,最差的分类方法比组合分类方法相差的正确率都超过了两个百分点,尤其在书籍领域,组合分类方法(0.811)比最差的分类方法(0.775)提高了 3.6%。因此,我们认为,为了保证系统分类效果达到最佳表现,组合分类方法是值得使用的。但是由于组合系统需要进行多个分类器的分类,系统在于测试阶段的时间复杂度要比使用单一分类方法要高。不过这种时间复杂度方面的提高是相对于分类器数目线性变化的,在实际应用中是可以接受的。

5 结论

在情感文本分类任务中,选择合适的分类方法直接影响系统的分类性能。为了避免不同的领域选择不同合适的分类方法,本文提出利用基于 Stacking 的组合分类方法组合四种不同性质的分类方法,分别为 NB、ME、SVM 和 SGD。在实验中,我们利用四个不同领域的中文情感文本分类语料测试了该组合分类方法。实验结果表明该方法能

够获得比固定融合算法(乘法规则)更好的分类结果,而且能够获得比基分类器最好结果更佳分类效果。

在下一步的研究工作中,我们将参考这些研究比较成熟的结果,利用情感文本更丰富的特征信息产生多样的分类器及如何实现动态的分类器选择方法等问题。例如,不同词性的特征对于整篇文章的情感分类的贡献不一样,我们可以组合不同词性特征产生的分类器,考察组合后的分类效果。另外,我们将重点分析情感(Sentiment)文本中感情(Emotion)表达的份额,并利用已有的感情分类语料构建一个独立于领域的情感分类器,用以融合到我们的系统中。

参考文献

- [1] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EM NLP-02). 2002.
- [2] 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类[J]. 中文信息学报, 2007, 21(6): 95-100.
- [3] 朱嫣岚, 闵锦, 周雅倩, 黄萱菁, 吴立德. 基于HowNet的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20.
- [4] 徐琳宏, 林鸿飞, 杨志豪. 基于语义理解的文本倾向性识别机制[J]. 中文信息学报, 2007, 21(1): 96-100.
- [5] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts [C] // Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL-04). 2004.
- [6] E. Riloff, S. Patwardhan, and J. Wiebe. Feature subsumption for opinion analysis [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EM NLP-06). 2006.
- [7] H. Cui, V. Mittal, and M. Datar. Comparative experiments on sentiment classification for online product reviews [C] // Proceedings of AAAI-06 the 21st National Conference on Artificial Intelligence. 2006.
- [8] 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报, 2007, 21(6): 88-94.
- [9] S. Tan and J. Zhang. An empirical study of sentiment analysis for Chinese documents [J]. Expert Systems with Applications. 2008, 34(4): 2622-2629.
- [10] J. Li and M. Sun. Experimental study on sentiment classification of Chinese review using machine learning techniques [C] // Processing of International Conference on Natural Language Processing and Knowledge Engineering, (NLP-KE-07), 2007.
- [11] M. Sahami. Learning limited dependence Bayesian classifiers [C] // Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996: 335-338.
- [12] V. Vapnik. The Nature of Statistical Learning Theory [M]. Springer, Berlin, 2005.
- [13] T. Zhang. Solving large scale linear prediction problems using stochastic gradient descent algorithms [C] // Proceedings of International Conference on Machine Learning (ICML-04). 2004.
- [14] J. Blitzer, M. Dredze, and F. Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain adaptation for sentiment classification [C] // Proceedings of Annual Meeting on Association for Computational Linguistics (ACL-07). 2007.
- [15] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20: 226-239.
- [16] R. Vilalta and Y. Drissi. A perspective view and survey of meta-learning [J]. Artificial Intelligence Review, 2002, 18(2): 77-95.
- [17] Saso Dzeroski and Bernard Zenko. Is combining classifiers with stacking better than selecting the best one? [J]. Machine Learning, 2004, 54(3): 255-273.
- [18] Rie Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data [J]. Journal of Machine Learning Research, 2005, 6: 1817-1853.