一种基于图的层次多标记文本分类方法

罗 俊
(广东技术师范学院 计算机与网络中心, 广州 510665)

摘 要: 由于一个类别在层次树上可能存在多个镜像,基于层次树来进行分类可能会导致不一致性。一种自然的解决方法是采用图结构来描述类别关系,在现实生活中人们实际的描述方式也是如此。鉴于此,提出了一种直接基于图的层次多标记分类方法,称为 GraphHMLTC。该方法利用有向无圈图的拓扑排序而非树的自顶向下的层次关系来确定类别之间的分类顺序,并且该拓扑序根据分类情形进行动态维护。实验表明,采用层次图分类的 GraphHMLTC 方法比非层次分类方法的代表之一 BoosTexter. MH 在较大程度上改善了分类精度。该工作体现了基于层次图的分类方法的可行性和优越性。

关键词: 文本分类; 层次分类; 多标记分类; 有向无圈图; 拓扑排序
中图分类号: TP181 **文献标志码**: A **文章编号**: 1001-3695(2010)03-0909-04
doi:10.3969/j. issn. 1001-3695. 2010. 03. 028

Graph-based method for hierarchical multi-lable text classification

LUO Jun
(*Center of Computer & Network*, *Guangdong Polytechnic Normal University*, *Guangzhou* 510665, *China*)

Abstract: Most of existing hierarchical text classification methods is based on a hierarchical category tree. However, such a tree structure maybe leads to some kinds of inconsistency for the reason of multiple images of a category on it. A nature solution for this is to adopt a hierarchical graph structure, which is a practical way to depict category relationships in a real world. So this paper presented a novel method for multi-lable text classification directly based on a hierarchical graph, called GraphHMLTC. Determined the classification order among categories by a topological sorting of vertexes in a graph (in fact, a directed acyclic graph), not by a hierarchical structure from top to down in a tree. Also, dynamically maintained the topological sorting according to the classification situation. Experiment results show that the method improves the classification accuracy in a great degree, compared to a representative of non-hierarchical multi-lable classification methods, BoosTexter. MH. Therefore, this work reveals that a graph-based classification method is feasible and superior.

Key words: text classification(TC); hierarchical classification; multi-lable classification; directed acyclic graph; topological sorting

近年来,文本分类(TC)的研究热点主要集中在数据集倾斜(imbalanced data set)、标注瓶颈(label bottleneck)、层次分类、问题的非线性可分性(nonlinear separability)以及 Web 页面分类(Web document categorization)等方面[1~3]。其中,层次分类是指多层类别关系下的分类问题,面对的类别间存在类似于树或有向非循环图的多层分级类别结构,可以更好地支持浏览和查询,也使得部分规模较大的分类问题通过分治的方法得到更好的解决。更进一步,如果允许一个文档的类别标记为层次类别结构中的 0、1 个或者多个类别,则称为层次多标记文本分类[4~6]。

随着基于学习的分类技术的进展,分类不一致性(classification inconsistency)问题日益受到人们的关注[7~9]。不一致性的现象多种多样,其根源主要来自于两个方面:分类方法本身和领域背景知识。前者如 SVM 层次分类方法[9],文档可能属于某个儿子类别但不属于其父亲类别;后者主要是体现在用来描述类别关系的层次结构上。到目前为止,几乎所有的层次文本分类方法[10~13]都采用树型结构来描述类别关系,即层次树(hierarchical tree)。层次树提供了一种简单且易于实现的描述方式,但是在类别较多且关系复杂的情况下会存在一些隐患。例如,经济法既属于经济范畴又属于法律范畴,这样的一个类别在层次树中既位于以经济类别为根的子树,同时也位于以法律为根的子树。因此,层次树最大的问题是面对同时属于多个类别的子类别,势必需要用不同的节点(即镜像)来表示它,这可能产生同一类别在不同子树中不一致的分类结果。

鉴于此,本文提出一种基于层次图(hierarchical graph)的多标记文本分类方法。该方法利用有向无圈图的拓扑排序而非树的自顶向下的层次关系来确定类别间的分类顺序。在层次图中每个类别只对应一个节点,可以避免层次树中一个类别对应多个节点可能导致的不一致;此外,层次图中顶点拓扑序的动态维护也可以防止如 SVM 方法所导致的不一致性。据笔者所知,目前还没有任何文本分类技术直接采用图结构来进行层次分类,因此本文的方法是很有研究前景和应用价值的。

1 层次文本分类简介

随着待分类的类别数目增多,类别本身固有的层次结构在分类过程中越发突显其重要性。层次文本分类的关键在于如

收稿日期: 2009-07-13; **修回日期**: 2009-08-21
作者简介:罗俊(1959-),男,浙江上虞人,副教授,硕士,主要研究方向为机器学习和网络技术(luo_jun_0523@163.com).

何正确而有效地利用类别关系来指引分类过程。基本方法分为两大类,即 big-bang 方法和 top-down 方法[10]。

在 big-bang 方法中,整个分类过程只使用一个分类器,将处于层次树上的所有叶节点类别看成平等的类,本质上是一种单层分类(flat classification),不能很好地应用类别间的关系。此外,分类器构造极其不灵活,一旦类别结构发生稍许变化就需要重新训练。在 top-down 方法中,层次树的每个节点都有一个分类器。文档首先由根节点的分类器进行分类,然后传递给下层的一个或多个儿子节点;再由儿子节点的分类器进行分类,直至到达叶子节点或某个不能再分类的中间节点。该方法的最大缺点是容易阻断问题(block problem)[8]。一旦父节点的分类器分类错误,文档将不能传递给儿子节点的分类器。此外,因为要构造很多分类器,所以需要充足的训练例,否则性能受影响。

在层次分类过程中,如果需要多标记分类则无疑增加分类任务的难度。层次的多标记分类即将实例标记为层次结构上的多个节点。多标记分类方法分为两大类,即问题转换方法和修改算法方法[4]。前者把多标记分类问题转换为一个或多个单标记分类或回归问题,后者扩展具体的学习算法来直接处理多标记数据。例如,Adaboost. MH 输出标记集合中属于每个标记的确认度[14];ML-kNN 对每个标记使用 KNN 算法然后输出标记的排序[5,15];概率模型生成方法[16]对每个标记产生不同的词;基于模型,多标记文档可以表示为标记的单词分布;模型的参数通过训练例最大化后验估计来学习,使用 EM 算法来计算哪些标记具有最大权值等。

2　层次图

2.1　层次结构分类

Sun 等人[10]将可采用的类别层次结构归纳为四种,即虚类别树(virtual category tree)、类别树(category tree)、虚有向无圈类别图(virtual directed acyclic category graph)和有向无圈类别图(directed acyclic category graph)。其中,在"虚"的层次结构中,文档只能标记为叶子节点关联的类别;而在非"虚"的层次结构中,文档可以标记为中间节点或者叶子节点关联的类别。

有向无圈类别图是现实生活中最常用的类别结构,如 Yahoo! 或 Open Directory Project 等构建的网络目录结构等。但是为了提高效率,目前已有的大多数方法都基于(虚)类别树,只有 Frommholz[7]尝试采用有向无圈类别图来进行层次分类,但是是在后处理阶段才考虑层次结构。非层次分类器先输出文档属于每个类别的概率,然后根据类别之间的相邻关系来调整这些概率值。由于分类过程依然是单层分类,降低了整体的分类性能。一些学习类别模型的方法[9]也将扩展到更通用的图结构上。

2.2　层次树的不足

当与某一类别相关联的类别较多时,就容易产生如前所述的在类别层次树上多个节点表示同一类别的现象,从而带来分类不一致的隐患。例如,图 1 是一棵对网页进行分类的层次树,根节点是一个虚节点,表示由所有网页文档组成的虚类别。其中的类别 A 和 B 既是类别 P_1 的子类,也是类别 P_2 的子类。按照 top-down 方法,对某篇文档 d 自顶向下进行分类,可能会产生如下结果:

a)P_1 的分类器认为 d 不属于 P_1,产生了阻断(即 d 也不属于 A);与此同时,P_2 的分类器认为 d 属于 P_2,继而往下传递,最后标记为 A。不同的子树对同一类别产生了不同的判断结果,这是第一种不一致现象。

b)假设分支节点的分类器最终是对叶子节点类别产生一个排序。在 P_1 的子树中,分类结果是 d 属于 B 的概率高于属于 A 的概率;而在 P_2 的子树中,分类结果是 d 属于 A 的概率高于属于 B 的概率。由于在不同子树中进行分类,很难对两个类别之间不同的排序结果进行统一处理,这是第二种不一致现象。

当然,还会存在由于树结构所导致的其他不一致现象。这些不一致现象的根源来自于树中不同节点可以表示同一类别,因此采用层次图可以避免这些问题。

2.3　层次图

层次图是有向无圈图,与层次树一样用来表示类别的种属关系。有向无圈图是不包含圈的有向图,在贝叶斯分类方法中也经常使用,用来表示特征之间的有限依赖关系[17]。

本文用来表示文本类别的层次图(图 2)有如下特点:a)每个节点表示一个类别,不同节点对应的类别也不同;b)每条有向边(即图 2 中的箭头线)表示类别之间的种属关系,箭尾表示父类,箭头指向子类;c)只有一个入度为 0 的节点,即表示由所有文档组成的总类;d)出度为 0 的节点表示不能再细分的类别。

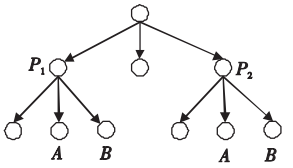

图1　分类网页文档的层次树

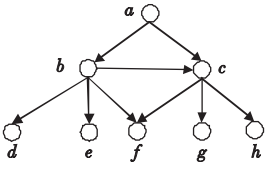

图2　类别层次图

不含平行边和自回路的图称为简单图。下面介绍后面章节用到的有向简单图中的相关概念和性质,来自于有向图的专著(文献[18])。

定义 1　设 $D=(V, A)$ 是有向简单图,V 是顶点集,A 是边集。对于 D 的一个顶点 v,定义集合:$N_D^+(v)=\{u\in V-v: vu\in A\}$,$N_D^-(v)=\{w\in V-v: wv\in A\}$,分别称集合 $N_D^+(v)$ 和 $N_D^-(v)$ 为顶点 v 的出邻集(out-neighbourhood)和入邻集(in-neighbourhood)。其中:$|N_D^+(v)|$ 称为 v 的出度 $\deg_+(v)$(out-degree),$|N_D^-(v)|$ 称为 v 的入度 $\deg_-(v)$(in-degree)。

例如,在图 2 的有向图中,$N_D^+(b)=\{c,d,e,f\}$,$N_D^-(b)=\{a\}$,且 $\deg_+(b)=4$,$\deg_-(b)=1$。

定理 1　有向无圈图一定存在着顶点的拓扑序。

拓扑序是图中顶点的排序,使得图中每条有向边的出发顶点在排序中都位于该边所指向的顶点的前面。Kahn[19]提出了复杂性为 $O(|V|+|A|)$ 的算法来找有向无圈图 $D=(V, A)$ 中的拓扑序。应用该算法在图 2 中可找到一条拓扑序为 $a\ b\ c\ d\ e\ f\ g\ h$。

3　算法

基于图的层次文本分类方法由两个过程组成(图 3),即训练过程和分类过程。训练过程使用训练文档来生成所需要的

分类器,而分类过程利用生成的分类器将新文档实例标记为合适的类别。由于层次图中顶点与类别是一一对应的,在不引起混淆的情况下,本文以下内容中将不再严格区分顶点与类别。

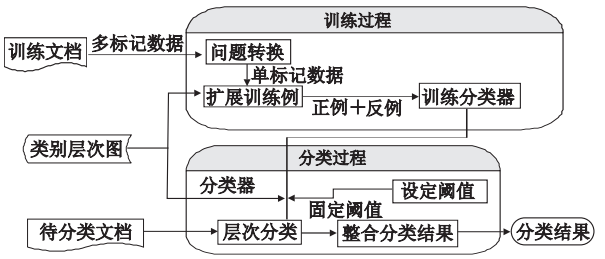

图3　基于层次图的分类算法框架

3.1　训练过程

由于一个文档可以属于多个类别,训练例是多标记数据(multi-lable data)。令 x 表示一个文档,Y 表示类别集合,则训练例的形式为 (x, Y^+)。其中:Y^+ 表示 x 所属的类别集合且 $Y^+ \subseteq Y$。因而在训练过程,首先要将训练例进行转换,使得多标记数据变为单标记数据(singlelable data),即一个训练例 (x, Y^+) 转换为 $|Y^+|$ 个训练例 (x, y^+)。其中 $y^+ \subseteq Y^+$。

另一方面,由于文档的标记过程一般只标记所属的类别,而不标记不属于的类别,训练集的反例如下:在层次图 D 中,设顶点 u 是顶点 v 的出邻集 $N_D^+(v)$ 的顶点,则类别 u 的反例集合 $Y^-(u)$ 为 $\left(\bigcup_{w \in (N_D^+(v)-u)} Y^+(w)\right) - Y^+(u)$,即一个类别的反例集等于位于同一出邻集的其他顶点的正例集合。

训练过程为层次图中的每个顶点 u(除了入度为 0 的总类别节点)生成一个二元分类器 $H_u: X \times Y \to R$。其中:X 是文档集合,□是类别集合,R 是实数集合。对于 $x \in X$ 且 $y \in Y$,$H_u(x, y) > 0$ 表示文档 x 属于类别 y,反之不属于。绝对值 $|H_u(x, y)|$ 则是对此判断的确认度。

3.2　分类过程

基于图的层次多标记文本分类过程见算法 GraphHMLTC。

Algorithm: Graph-based hierarchical multi-label text classification (GraphHMLTC)

输入:新文档 x,层次图 $HG = \langle V, E \rangle$,固定阈值 θ,类别集合 L。

输出:x 所属的每个类别 l 及其确认度 $H_l(x)$。

a)找出图 HG 中的一条拓扑序 T。

b)对 HG 中每个顶点 u 计算其入度 $\deg_-(u)$。

c)按照拓扑序 T 依次处理图 HG 中的顶点 u。

d)调用顶点 u(对应类别 l)的分类器 H_l 对 x 进行分类,返回结果 $H_l(x)$。

e)如果 $H_l(x) < \theta$,则从拓扑序 T 中去除 u,并且所有属于 u 的出邻集中顶点 v 的 $\deg_-(v)$ 减 1;如果 $\deg_-(v)$ 等于 0,则从拓扑序 T 中去除 v,且属于 v 的出邻集中的所有顶点入度减 1。依此类推,直到去除一个顶点后,在 T 中排在该顶点后面的所有顶点的入度均不为 0 为止。

f)输出 T 中保留的顶点所对应的类别 l 及其确认度 $H_l(x)$。

关于该算法的相关解释如下:

a)在层次图顶点的拓扑序中,父类所对应的顶点一定排在子类所对应的顶点前面。根据拓扑序来对文本进行分类(步骤 c)),可以保证先对父类进行分类再对子类进行分类。

b)在二元分类中,对文档属于某个类别的判别往往采用固定阈值或动态阈值的方法[20]。固定阈值是预先指定的最小确认度,而动态阈值是当前判别的类别集合的确认度平均值。固定阈值方法缺乏灵活性,但是稳定并且可以较好地表达对分类结果的期望;动态阈值方法灵活,但是在确认度都较小甚至为负值的情况下也要被迫选择文档归属的类别。基于上述分析,本文采用固定阈值方法(步骤 e))。

c)当一个分类器的返回结果没有达到固定阈值(返回结果为负值,表示文档不属于该类别;返回结果为正值,但是小于固定阈值,表示文档属于该类别的确认度不高),即表示文档不属于该类别。按照层次分类的思想,如果文档不属于某一父类,则也不属于其子类,因而不需要调用子类的分类器进行分类。故算法的步骤 e)根据分类情形对拓扑序进行动态调整,剪除已经明确不属于某父类对其子类的传递路径(即入度减 1)。如果某类别顶点的入度减为 0,则表示待分类文档不属于其所有父类,因此需从拓扑序中去除该顶点,即无须调用该类别的分类器。

d)拓扑序中没有被去除的顶点对应的类别即待分类文档所属的类别(步骤 f))。

由于找拓扑序的时间为 $O(|V| + |E|)$,算法 GraphHMLTC 的时间复杂性为 $\max\{O(|V| + |E|), O(|V|T(H_l))\}$。其中 $T(H_l)$ 为二元分类器 H_l 所花的时间。

关于拓扑序的动态调整举例如下:图 2 中顶点的原有拓扑序为 $a\ b\ c\ d\ e\ f\ g\ h$,设 $\theta = 0.5$,如果对于文档 x,有 $H_b(x) = 0.3 < \theta$,则 $\deg_-(d) = \deg_-(e) = 0$,$\deg_-(f) = 1$,因此拓扑序变为 $a\ c\ f\ g\ h$。

4　实验

为了研究本文方法的可行性,选取了 Open Directory Project 上的一个子集 computers 来进行实验。该子集的部分类别层次图如图 4 所示。实验过程共使用了该子集下的 381 个类别和 6 953 个网页文档。Computers 类别分为 6 个子类,各子类所选择的文档数目如表 1 所示。

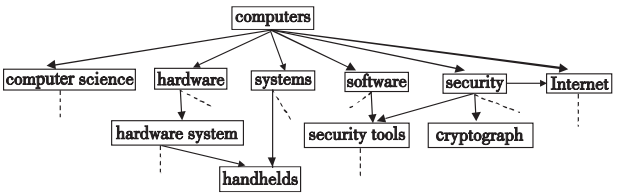

图4　Open Directory Project的computers子集层次图

表 1　computers 的 6 个子类的文档数

类别	文档数	类别	文档数
computer science	187	security	275
hardware	568	software	2906
Internet	2 713	system	304

本文所选用的评价文本分类性能的指标包括查全率(recall)、查准率(precision)、宏观 F_1 值(macro-F_1)、微观 F_1 值(micro-F_1)。设 N_{CR_i} 是正确分类到 C_i 类的文档数,N_{C_i} 是实际属于 C_i 类的文档数,N_{P_i} 是分类器预测为 C_i 类的文档数,N 是文档总数,m 是类别总数。各指标的具体定义如下:

查准率:$P_i = N_{CR_i} / N_{P_i}$

查全率:$R_i = N_{CR_i} / N_{C_i}$

宏观查准率:$P^M = (1/m)\sum_{i=1}^{m} P_i$

宏观查全率:$R^M = (1/m)\sum_{i=1}^{m} R_i$

F_1 值:$F_{1i} = 2R_iP_i/(R_i + P_i)$

宏观 F_1 值:$F_1^M = (1/m)\sum_{i=1}^{m} F_{1i}$

微观 F_1 值：$F_1^{\mu}=(2\times\sum_{i=1}^{m}P_i\times\sum_{i=1}^{m}R_i)/((\sum_{i=1}^{m}P_i+\sum_{i=1}^{m}R_i)\times m)$

本文选择非层次分类方法的代表——BoosTexter. MH[14]与本文方法作比较。BoosTexter. MH 是一个单层的多标记文本分类器，使用 Boosting 技术来改善分类精度。为了方便实现，本文算法 GraphHMLTC 的层次图上每个节点的二元分类器均使用 BoosTexter. MH(即单层单标记分类)，利用节点“本地的”训练例来判别是否属于对应的类别。而在 BoosTexter. MH 的实验中，所有类别的文档作为一个完整的训练集来进行。

实验环境为 Pentium Processor(2.66 GB) + RAM(1 GB) + Fedora 9.0。实验过程采用 3 交叉验证法，即训练集分为三个子集，两个用来训练，一个用来验证。实验结果如表 2 所示。其中 GraphHMLTC 的阈值设为 $\theta=0.5$。表 2 的第一行表示实验中所采用的 Boosting 技术的迭代次数 T，分别为 5、10、20、50、100、200 次。每栏中的四个数字从上到下依次为 P^M、R^M、F_1^M、F_1^{μ} 的值。例如，当迭代次数 $T=5$ 时，在训练集上 GraphHMLTC 的宏观查准率 $P^M=0.54$，宏观查全率 $R^M=0.56$，宏观 F_1 值 $F_1^M=0.55$，微观 F_1 值 $F_1^{\mu}=0.50$。

表 2　本文方法与 BoosTexter. MH 的分类性能比较

分类方法	5	10	20	50	100	200
BoosTexter. MH	0.43	0.45	0.50	0.58	0.63	0.66
	0.51	0.48	0.56	0.57	0.68	0.68
	0.47	0.46	0.53	0.57	0.65	0.70
	0.42	0.43	0.50	0.56	0.61	0.65
GraphHMLTC	0.54	0.59	0.65	0.74	0.82	0.84
	0.56	0.56	0.67	0.70	0.80	0.79
	0.55	0.57	0.66	0.71	0.81	0.81
	0.50	0.51	0.60	0.68	0.77	0.79

从表 2 可以看到，GraphHMLTC 的性能是令人满意的。随着迭代次数的增多，两种方法的宏观查准率 P^M 和宏观查全率 R^M 都有所提高，但是 GraphHMLTC 增长的幅度要比 BoosTexter. MH 大得多。特别地，在迭代次数为 200 时，GraphHMLTC 的 P^M 已经达到 84%，而 BoosTexter. MH 仅有 66%，这说明采用图形层次结构在较大程度上提高了分类精度。另外，GraphHMLTC 的 F_1^M 和 F_1^{μ} 也呈稳定增长的趋势，这说明由查准率和查全率所构成的综合性能指标也改善了。

5　结束语

本文提出了一种基于图结构的层次多标记分类方法，称为 GraphHMLTC。与基于树结构的层次分类方法不同，该方法利用有向无圈图的拓扑排序来确定类别之间的分类顺序。层次图中的每个类别只对应一个节点，在其之上的拓扑排序保障了先父类后子类的合理分类顺序。而使用图结构就自然克服了树结构中由于多个节点表示同一类别所带来的分类不一致性。实验表明，采用层次图分类的 GraphHMLTC 方法比非层次分类方法 BoosTexter. MH 在较大程度上改善了分类的精度。

本文的未来工作如下：a)进一步完善实验工作。由于大部分基于树结构的层次分类器不能直接获得，但可以尝试实现其近似版本，来与本文方法进行分类性能和代价等方面的比较。b)增强方法的应用性。将基于图的层次分类思想应用到更多的多标记分类领域以提高准确度，如音乐分类、语义情景分类、医学诊断等领域。

参考文献：

[1] 苏金树，张博锋，徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报，2006，17(9)：1848-1859.

[2] 郝秀兰，陶晓鹏，徐和祥，等. KNN 文本分类器类偏斜问题的一种处理对策[J]. 计算机研究与发展，2009，46(1)：52-61.

[3] 周炎涛，唐剑波，吴正国. 基于向量空间模型的多主题 Web 文本分类方法[J]. 计算机应用研究，2008，25(1)：142-144.

[4] TSOUMAKAS G, KATAKIS I. Multi-label classification：an overview [J]. International Journal of Data Warehousing and Mining, 2007，3(3)：1-13.

[5] TSOUMAKAS G, VLAHAVAS I. Random k-labelsets：an ensemble method for multilabel classification[C]//Proc of the 18th European Conference on Machine Learning：Springer, 2007：406-417.

[6] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data[K]//Data Mining and Knowledge Discovery Handbook. 2nd ed. New York：Springer, 2009：1383.

[7] FROMMHOLZ I. Categorizing Web documents in hierarchical catalogues[C]//Proc of the 23rd European Colloquium on Information Retrieval Research. Darmstadt, Delaware：Springer, 2001.

[8] SUN A, LIM E P, NG W K, *et al*. Blocking reduction strategies in hierarchical text classification[J]. IEEE Trans on Knowledge and Data Engineering, 2004，16(10)：1305-1308.

[9] ROUSU J, SAUNDERS C, SZEDMÁK S, *et al*. Learning hierarchical multi-category text classification models[C]//Proc of the 22nd International Conference on Machine Learning. New York：ACM Press, 2005：744-751.

[10] SUN Ai-xin, LIM E P. Hierarchical text classification and evaluation [C]//Proc of IEEE International Conference on Data Mining. Washington DC：IEEE Computer Society, 2001：521-528.

[11] SUN Ai-xin, LIM E P. Web unit based mining of homepage relationships[J]. Journal of the American Society for Information Science and Technology, 2006，57(3)：394-407.

[12] HUANG C C, CHUANG S L, CHIEN L F. Liveclassifier：creating hierarchical text classifiers through Web corpora[C]//Proc of the 13th International ACM World Wide Web Conference. New York：ACM Press, 2004：184-192.

[13] ESULI A, FAGNI T, SEBASTIANI F. TreeBoost. MH：a boosting algorithm for multi-label hierarchical text categorization[C]//Proc of the 13th International Symposium on String Processing and Information Retrieval. Berlin：Springer, 2006：13-24.

[14] SCHAPIRE R E, SINGER Y. BoosTexter：a boosting-based system for text categorization[J]. Machine Learning, 2000，39(2/3)：135-168.

[15] ZHANG Ming-lin, ZHOU Zhi-hua. A k-nearest neighbor based algorithm for multi-label classification[C]//Proc of the 1st IEEE International Conference on Granular Computing. 2005：718-721.

[16] McCALLUM A. Multi-label text classification with a mixture model trained by EM[C]//Proc of the AAAI Workshop on Text Learning. Orlando, Florida：AAAI, 1999.

[17] KOLLER D, SAHAMI M. Hierarchically classifying documents using very few words[C]//Proc of the 14th International Conference on Machine Learning. San Francisco：Morgan Kaufmann, 1997：170-178.

[18] BANG-JENSEN J, GUTIN G D. Theory, algorithms and applications [M]//2nd ed. London：Springer, 2008.

[19] KAHN A B. Topological sorting of large networks[J]. Communications of the ACM, 1962，5(11)：558-562.

[20] 吴春颖，王士同. 一种改进的 KNN Web 文本分类方法[J]. 计算机应用研究，2008，25(11)：3275-3277.