# 基于局部近邻相关性的多标记算法

郑希源 张化祥
(山东师范大学信息科学与工程学院 济南 250014)
(山东省分布式计算机软件新技术重点实验室 济南 250014)

**摘 要** 通过近邻样例类标记确定测试样例类标记的思想在多标记分类算法中取得了良好的效果。该类算法通过对训练集进行学习,建立训练样例类标记与其 $k$ 个近邻样例中不同类标记样例个数的映射关系,然后用该映射关系预测测试样例的类标记。该类算法的不足是只考虑近邻样例中不同类别样例的个数与测试样例类标记的映射关系,忽略了近邻样例与测试样例的局部相关性。考虑训练样例类与近邻样例的局部相关性,建立起它们类别间的映射关系,预测测试样例类标记,提出 ML-WKNN 算法。实验表明,ML-WKNN 能更好地处理多标记分类问题和自动图像标注问题。

**关键词** 多标记学习,$k$ 近邻,分类,局部相关
**中图法分类号** TP181 **文献标识码** A

## Multiple Label Approach Based on Local Correlation of Neighbors

ZHENG Xi-yuan ZHANG Hua-xiang
(Department of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)
(Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Jinan 250014, China)

**Abstract** Determining the classification of the test sample by using neighbors' labels achieves good results in multiple label classification. The mapping relationships of these algorithms are established between the labels of training examples and the number of different samples in their $k$-nearest neighbors by learning from the training set. The label of a test sample can be easily predicted by applying the mapping relationship. The disadvantage of these algorithms is to consider only the mapping relationship between the labels of the test examples and the number of different samples in their $k$-nearest neighbors, and to ignore the local correlation between the labels of the test examples and their $k$-nearest neighbors. This paper proposed an algorithm called ML-WKNN algorithm, which classifies the test examples through the mapping relationship between the labels of the training examples and their $k$-nearest neighbors by considering the local correlation between the labels of the training examples and their $k$-nearest neighbors. The experimental results show that the ML-WKNN algorithm achieves better results than other algorithms in dealing with the multi-label classification problems and automatic image annotation.

**Keywords** Multi-label learning, KNN, Classification, Local correlation

## 1 引言

多标记学习已成为机器学习领域的研究热点之一[1-3]。与传统单标记学习相比,多标记学习能更好地描述现实世界中的事物和现象。单标记学习中一个示例只对应一个标记,而实际上示例通常有多个语义,如文本分类中,一个文本可能同时属于多个主题[1,2]。网页分类中,含有贝克汉姆图像的网页可被归为体育新闻类,同时也可归为娱乐、旅游、经济类等;图像分类中,含有多个对象的图像可依据图像内容归到不同类别[4]。

常见多标记学习算法有问题转化法和算法适应法[5,6]。其中问题转化法首先使用有关规则将多标记问题转换成一个或几个单标记问题,使用传统单标记学习技术处理转化后的问题。方法简单易实现,但在类标记间存在相互关系的多标记问题上表现不好[1-4]。算法适应法将单标记学习算法做某种调整和加强,处理多标记问题,主要有:用于文本分类的多标记算法,基于决策树的多标记学习算法[7,8],基于核函数的多标记学习算法[4,9],基于神经网络的多标记学习算法[10]和结合贝叶斯理论与 k 近邻的多标记学习算法 ML-KNN[5]。

上述有关算法中,ML-KNN 算法在多标记分类问题上性能相对更好,算法不足之处是只考虑近邻样例中不同类别样例的个数与测试样例类标记的映射关系,忽略了近邻样例与测试样例距离上的相关性。本文提出的局部近邻相关多标记算法(ML-WKNN),从训练样例类标记与近邻样例中不同类别样例的欧式距离和个数两个方面考虑,建立起它们间的映射关系,然后用该映射关系预测测试样例类标记。

到稿日期:2013-05-20 返修日期:2013-07-23 本文受国家自然科学基金(61170145),教育部高等学校博士点专项基金(20113704110001),山东省自然科学基金和科技攻关计划项目(ZR2010FM021,2010G0020115)资助。
**郑希源**(1979—),男,博士生,主要研究方向为机器学习、信息检索及数据挖掘等,E-mail:9919005@163.com;**张化祥**(1966—),男,博士,教授,博士生导师,主要研究方向为机器学习、模式识别及 Web 挖掘等,E-mail:huaxzhang@163.com(通信作者)。

本文第 2 节介绍 ML-KNN 算法；第 3 节描述 ML-WKNN 算法；第 4 节介绍评测标准；第 5 节给出实验数据集并分析实验结果；最后对本文进行总结。

## 2 ML-KNN 算法[5]

ML-KNN 算法综合 $k$ 近邻(KNN)和贝叶斯算法处理多标记学习问题。KNN 通过测试样例的 $k$ 个近邻决定测试样例的类别。贝叶斯算法从给定训练数据集中得到类标记先验概率和条件概率，然后利用贝叶斯法则计算测试样例最大后验概率，确定测试样例类别[11,12]。ML-KNN 不同于 KNN，该算法不是简单地通过 $k$ 个近邻样例投票确定测试样例类别，它同时从训练集中学习训练样例的近邻样例中不同类标记样例个数和训练样例类标记间的概率关系，比只考虑测试样例近邻样例中不同类标记样例个数更合理全面，因为利用概率分类方法可有效克服噪声数据干扰。

ML-KNN 算法思想如下：

训练集 $D=\{(x_i,Y_i)\mid 1\leqslant i\leqslant\psi\}$，$Y\subseteq\Omega$，$\Omega$ 是类标记集合，样例 $x$ 的类标记向量 $\overrightarrow{y_x}$ 的第 $\iota$ 个类标记 $\overrightarrow{y_x}(\iota)=1$ 表示样例 $x$ 具有类标记 $\iota$，$\overrightarrow{y_x}(\iota)=0$ 表示样例 $x$ 不具有类标记 $\iota$，$N(x)$ 表示样例 $x$ 在训练集中的 $k$ 近邻集合，样例 $x$ 的 $k$ 个近邻中具有类标记 $\iota$ 的样例数量由式(1)计算：

$$\overrightarrow{C_x}(\iota)=\sum_{a\in N(x)}\overrightarrow{y_a}(\iota),\iota\in\Omega \tag{1}$$

$N(t)$ 表示测试样例 $t$ 在训练集中的 $k$ 个近邻集合，$H_1^\iota$ 表示测试样例 $t$ 具有类标记 $\iota$，$H_0^\iota$ 表示测试样例 $t$ 不具有类标记 $\iota$，$E_j^\iota$ 表示 $N(t)$ 中恰好有 $j$ 个样例具有类标记，测试样例 $t$ 的类标记与 $N(t)$ 中具有该标记或者不具有该标记的样例数量的最大后验概率关系如下：

$$\overrightarrow{y_t}(\iota)=\arg\max_{b\in\{0,1\}}P(H_b^\iota\mid E^\iota_{\overrightarrow{C_t}(\iota)}) \tag{2}$$

根据贝叶斯法则，式(2)可改写为：

$$\overrightarrow{y_t}(\iota)=\arg\max_{b\in\{0,1\}}\frac{P(H_b^\iota)P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota)}{P(E^\iota_{\overrightarrow{C_t}(\iota)})} \tag{3}$$

$$=\arg\max_{b\in\{0,1\}}P(H_b^\iota)P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota) \tag{4}$$

由式(4)可知，为确定测试样例类向量中各元素的值，必须首先依据训练样例计算所对应类标记的先验概率 $P(H_b^\iota)(\iota\in\Omega,b\in\{0,1\})$ 和条件概率 $P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota)(j\in\{0,1,\cdots,k\})$。

## 3 局部近邻相关的多标记算法 ML-WKNN

ML-KNN 只考虑近邻样例中不同类别样例的个数与测试样例类标记的映射关系，忽略了近邻样例与测试样例距离上的相关性。如当测试样例 $t$ 的 $k$ 个近邻中某类样例在数量上占优时，该测试样例被确定为该类样例却不一定正确。如果 $k$ 近邻中该类样例距离测试样例远，其它样例距离测试样例近，那么该测试样例与个数少但与其距离近的样例划为一类可能是正确的。如图 1 所示(十字点、小三角代表两类样例，空心点代表测试样例)。

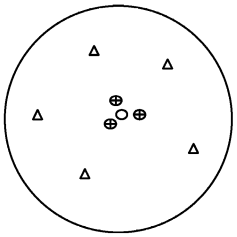

图 1 $k$ 近邻示意图

基于上述分析，本文提出局部近邻相关的多标记算法，从训练样例类标记与近邻样例中不同类别样例的欧式距离和个数两方面的相关性考虑，建立训练样例类标记与近邻样例中不同类别样例的欧式距离和个数间的映射关系，用该映射关系预测测试样例的类标记。把训练样例与其近邻间距离平方的倒数看作近邻样例对测试样例分类影响力的权重 $W_b^\iota$。ML-WKNN 算法中确定测试样例类别的最大后验概率计算公式如下：

$$\overrightarrow{y_t}(\iota)=\arg\max_{b\in\{0,1\}}P(H_b^\iota)P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota)W_b^\iota \tag{5}$$

式(5)中 $P(H_b^\iota)$、$P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota)$ 与式(4)中意义相同，式中 $W_b^\iota(b\in\{0,1\})$ 的值从训练样例集中学习得到，$W_1^\iota$ 表示当某个训练样例具有类标记 $\iota$ 时，其近邻样例中具有该类标记样例的权重之和与不具有该类标记样例权重之和的比值；$W_0^\iota$ 表示某个训练样例不具有类标记 $\iota$ 时，其 $k$ 个近邻样例中具有该类标记样例的权重之和与不具有该类标记样例权重之和的比值。

ML-WKNN 算法描述：

(1)从训练集中计算出类标记集合中各类标记的先验概率 $P(H_b^\iota)(\iota\in\Omega,b\in\{0,1\})$；

(2)计算训练集中任意两样例的欧式距离，得到距离矩阵，并根据距离从小到大排序；

(3)对任意训练样例 $x_i$，根据欧式距离取出其 $k$ 个近邻放入集合 $N(x)$，取出 $x_i$ 与 $k$ 个近邻的欧氏距离放入距离矩阵 $D_{is}$ 中；

(4)对于类标记集合中的任意类标记，求出任意训练样例 $x_i$ 的 $k$ 个近邻中含有该类标记的近邻数量；

(5)计算任意训练样例 $x_i$ 及其 $k$ 个近邻中关于该类标记的情况；

(6)统计得到任意训练样例 $x_i$ 及其 $k$ 个近邻中关于该类标记的分布规律；

(7)计算当任意训练样例 $x_i$ 具有某个类标记时，$x_i$ 及其 $k$ 个近邻中具有该类标记的样例数量为多少时概率最大，当任意训练样例 $x_i$ 不具有某个类标记时，$x_i$ 及其 $k$ 个近邻中不具有该类标记的样例数量为多少时概率最大；

(8)对于任意训练样例 $x_i$，取该样例与其 $k$ 个近邻的欧氏距离的平方的倒数作为每个近邻的权重，然后归一化，得到权重矩阵；

(9)当 $x_i$ 具有某个类标记时，计算其 $k$ 个近邻中具有该类标记近邻的权重之和与不具有该类标记近邻权重之和的比值，并统计出最常见比值概率 $P_w=W_1^\iota$，同样计算当 $x_i$ 不具有某个类标记时，其 $k$ 个近邻中不具有该类标记近邻的权重之和与具有该类标记近邻权重之和的比值，并统计出最常见比值概率 $P_{wn}=W_0^\iota$；

(10)对一个测试样例进行分类时，首先在训练集中找出该测试样例的 $k$ 个近邻，把从训练样例集中学习得到的先验概率 $P(H_b^\iota)(\iota\in\Omega,b\in\{0,1\})$、条件概率 $P(E^\iota_{\overrightarrow{C_t}(\iota)}\mid H_b^\iota)(j\in\{0,1,\cdots,k\})$ 及 $W_b^\iota(b\in\{0,1\})$ 代入式(5)，计算确定给定测试样例的分类。

## 4 评测标准

本文应用下列评测标准对文中多标记学习算法进行评测[2]：

(1)Average precision

该指标用于计算排名高于某一特定标签的平均分数，取值范围从0到1，取值等于0时，说明算法性能最差，取值越大，算法性能越好。公式表示如下：

$$avgprec_S(f)=\frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|}*\frac{|\{y'|rank_f(x_i,y')\leqslant rank_f(x_i,y),y'\in Y\}|}{rank_f(x_i,y)} \tag{6}$$

(2)Coverage

该指标用来说明为覆盖训练样例的所有标记需要在样本类标记序列中执行深度的情况，取值范围从0到1。取值等于0时，算法性能最佳，取值越大，算法性能越差。公式表示如下：

$$coverage_S(f)=\frac{1}{p}\sum_{i=1}^{p}\max_{y\in Y_i}rank_f(x_i,y)-1 \tag{7}$$

(3)Hamming loss

该指标用来说明示例-标记对被错分的次数，即某个示例有某个标记却没有被预测，某个示例本该没有某个标记却被预测了，取值范围从0到1。取值等于0时，算法性能最佳，取值越大，算法性能越差。公式表示如下：

$$hloss_S(h)=\frac{1}{p}\sum_{i=1}^{p}\frac{1}{Q}|h(x_i)\Delta Y_i| \tag{8}$$

(4)One-error

该指标用来说明不属于某个示例标记集中的标记却在预测中排名第一的次数，取值范围从0到1。取值等于0时，算法性能最佳，取值越大，算法性能越差。公式表示如下：

$$one\text{-}error_S(f)=\frac{1}{p}\sum_{i=1}^{p}[[\arg\max_{y\in Y_i}f(x_i,y)]]\notin Y_i] \tag{9}$$

(5)Ranking loss

该指标用来说明对于某个样例的标记错误排序的情况，取值范围从0到1。取值等于0时，算法性能达到最佳，取值越大，算法性能越差。公式表示如下：

$$rloss_S(f)=\frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i\|\overline{Y}_i|}|\{(y_1,y_2)|f(x_i,y_1)\leqslant f(x_i,y_2),(y_1,y_2)\in Y_i\times\overline{Y}_i|\} \tag{10}$$

## 5 实验结果

将本文算法与当前较流行的4个算法从Hamming loss、One-error、Coverage、Ranking loss、Average precision 5个方面进行分析比较，它们分别是基于Boosting思想的BoosTexter算法、基于决策树思想的AdaBoost.MH算法、基于核函数思想的RANK-SVM算法及基于贝叶斯理论和近邻思想的ML-KNN。

本文在一个生物信息学数据和一个自然场景分类的数据集上对有关算法进行评价。

### 5.1 Yeast基因功能分析

本文首先通过预测Yeast酿酒酵母菌的基因功能分类来评价我们提出的算法。为了便于操作，本文实验使用与文献[5]相同的数据内容。所用数据集共有2417个样例，每个样例均有103个属性值和14个可选类标记。每个样例平均类标记个数为4.24±1.57。

本文采用十折交叉验证方法进行实验。实验结果如表1所列。本文算法在$k$取5至9时，5个评测指标取得最佳值。ML-KNN算法在$k$取8至12时，5个评测指标相对较好。

表1 ML-WKNN算法在Yeast数据集的实验结果

| 评价标准 | 不同k值 | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| Average precision | 0.7581 | 0.7576 | **0.7599** | 0.7564 | 0.7551 |
| Coverage | 6.4000 | 6.3922 | **6.3620** | 6.3770 | 6.4089 |
| Hamming loss | **0.1965** | 0.1968 | 0.1972 | 0.1968 | 0.1987 |
| One-error | 0.2454 | 0.2377 | **0.2366** | 0.2410 | 0.2519 |
| Ranking loss | 0.1711 | 0.1713 | **0.1701** | 0.1717 | 0.1731 |

表1显示了5个评价指标与不同$k$值之间的对应关系，表2显示了ML-KNN算法与不同$k$值之间的对应关系。在表1和表2中用黑体字标出ML-WKNN与ML-KNN算法的最佳实验结果。对比表1和表2中黑体数据可以发现：首先，在Yeast数据集上，ML-WKNN算法最好指标中有4项明显优于ML-KNN算法，1项实验结果略低于ML-KNN算法。其次，当$k=7$时，ML-WKNN算法有3个评测指标优于ML-KNN算法的最佳值，Hamming loss指标相等。第三，ML-WKNN算法在5个评测指标中的最好结果有4项在$k=7$时取得，有1项在$k$值为5时取得，ML-KNN算法在5个评测指标中的最好结果分别在$k=8$、$k=10$和$k=12$时取得，ML-WKNN算法的最好结果集中程度明显高于ML-KNN算法。

表2 ML-KNN算法在Yeast数据集的实验结果

| 评价标准 | 不同k值 | | | | |
|---|---|---|---|---|---|
| | 8 | 9 | 10 | 11 | 12 |
| Average precision | 0.7558 | 0.7553 | **0.7578** | 0.7572 | 0.7568 |
| Coverage | **6.3921** | 6.4351 | 6.4144 | 6.4000 | 6.4089 |
| Hamming loss | 0.1978 | 0.1974 | 0.1980 | 0.1984 | **0.1972** |
| One-error | 0.2475 | 0.2530 | 0.2345 | 0.2410 | **0.2339** |
| Ranking loss | **0.1716** | 0.1726 | 0.1719 | 0.1724 | 0.1726 |

表3给出ML-WKNN算法与其它4种多标记分类算法最佳实验结果的对比情况。可以看出，本文算法在4项评测指标方面优于其它4个算法，只有1项指标略低于ML-KNN算法，但也明显优于另外3种算法。

表3 ML-WKNN算法在Yeast数据集与其它算法的对比

| 评价标准 | 算法 | | | | |
|---|---|---|---|---|---|
| | ML-WKNN | ML-KNN | RANK-SVM | AdaBoost.MH | Boos Texter |
| Average precision | **0.7599** | 0.7578 | 0.7391 | 0.7382 | 0.7360 |
| Coverage | **6.3620** | 6.3921 | 7.5392 | 6.5672 | 6.5534 |
| Hamming loss | **0.1965** | 0.1972 | 0.2017 | 0.2012 | 0.2301 |
| One-error | 0.2366 | **0.2339** | 0.2510 | 0.2516 | 0.2797 |
| Ranking loss | **0.1701** | 0.1716 | 0.1951 | N/A | 0.1867 |

### 5.2 自动图像标注(Automatic Image Annotation)

图像标注和自然场景的多标记分类是等同的，因为多标记分类中的每一个标记可以看成是用于图像标注的关键字。本文使用与文献[5]相同的图像数据集评价我们提出的ML-WKNN算法。该数据集共有2000幅图像，每幅图像都手工标注了类标记，图像的可选标注词有沙漠、高山、海洋、落日和树木。图像的类标记统计情况如表4所列。数据集中22%以上的图像有1个以上的标注词，其中多数图像具有2个标注词，具有3个标注词的图像非常少，数据集中的每幅图像平均具有1.24个标注词。图像的特征向量表示方法与文献[13]相同。首先把图像导入到CIE Luv色彩空间，在该空间中图像的表示数据适合用欧式距离方法计算图像之间的差异，经过一系列处理之后，每幅图像用294维的特征向量来表示，详细情况见表4。

表 4　自然场景图像标记情况

| 类标记 | 图像数量 |
|---|---|
| Desert | 340 |
| Mountains | 268 |
| Sea | 341 |
| Sunset | 216 |
| Trees | 378 |
| Desert+mountains | 19 |
| Desert+sea | 5 |
| Desert+sunset | 21 |
| Desert+trees | 20 |
| Mountains+sea | 38 |
| Mountains+sunset | 19 |
| Mountains+trees | 106 |
| Sea+sunset | 172 |
| Sea+trees | 14 |
| Sunset+trees | 28 |
| Desert+mountains+sunset | 1 |
| Desert+sunset+trees | 3 |
| Mountains+sea+trees | 6 |
| Mountains+sunset+trees | 1 |
| Sea+sunset+trees | 4 |
| Total | 2000 |

在此数据集上，我们也用十折交叉验证的方法进行实验。实验结果如表 5 所列。

表 5　ML-WKNN 算法在 Image 数据集的实验结果

| 评价标准 | 不同 k 值 | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| Average precision | 0.7679 | **0.7898** | 0.7885 | 0.7839 | 0.7726 |
| Coverage | 0.9750 | 0.8650 | **0.8550** | 0.8650 | 0.9050 |
| Hamming loss | 0.1610 | 0.1630 | **0.1430** | 0.1620 | 0.1450 |
| One-error | 0.3300 | **0.3100** | 0.3150 | 0.3250 | 0.3400 |
| Ranking loss | 0.2096 | 0.1863 | **0.1838** | 0.1850 | 0.1938 |

在 Image 数据集中 ML-WKNN 算法也是在 $k$ 取 5 至 9 时，5 个评测指标取得最佳值。ML-KNN 算法的 $k$ 值取 8 至 12 时，5 个评测指标相对较好。表 5 显示了 5 个评价指标与不同 $k$ 值之间的对应关系，表 6 显示了 ML-KNN 算法与不同 $k$ 值之间的对应关系。在表 5 和表 6 中用黑体字标出 ML-WKNN 与 ML-KNN 算法的最佳实验结果。对比表 5 和表 6 中黑体数据可以发现，在 Image 数据集上，ML-WKNN 算法最好指标中有 4 项明显优于 ML-KNN 算法，仅 1 项实验结果略低于 ML-KNN 算法。

表 6　ML-KNN 算法在 Image 数据集的实验结果

| 评价标准 | 不同 k 值 | | | | |
|---|---|---|---|---|---|
| | 8 | 9 | 10 | 11 | 12 |
| Average precision | 0.7268 | 0.7389 | **0.7578** | 0.7510 | 0.7375 |
| Coverage | 1.1450 | 1.1650 | **1.1050** | 1.1450 | 1.2050 |
| Hamming loss | 0.1800 | 0.1810 | 0.1780 | **0.1720** | 0.2000 |
| One-error | 0.4000 | 0.4000 | **0.3650** | 0.3850 | 0.4250 |
| Ranking loss | 0.1980 | 0.1879 | **0.1716** | 0.1867 | 0.2013 |

表 7　ML-WKNN 算法在 Image 数据集与其它算法的对比

| 评价标准 | 算法 | | | | |
|---|---|---|---|---|---|
| | ML-WKNN | ML-KNN | RANK-SVM | AdaBoost.MH | Boos Texter |
| Average precision | **0.7898** | 0.7578 | 0.6791 | 0.7360 | 0.7502 |
| Coverage | **0.8550** | 1.1050 | 1.3920 | 1.2056 | 1.1037 |
| Hamming loss | **0.1430** | 0.1720 | 0.2800 | 0.2012 | 0.1805 |
| One-error | **0.3100** | 0.3650 | 0.5350 | 0.3706 | 0.3600 |
| Ranking loss | 0.1838 | **0.1716** | 0.2790 | N/A | 0.1717 |

表 7 给出 ML-WKNN 算法与其它 4 种多标记分类算法最佳实验结果的对比情况。可以看出，本文算法在 4 项评测指标方面优于其它 4 种算法，只有 1 项指标略低于 ML-KNN 算法和 BoosTexter 算法，但也明显优于另外两种算法。

**结束语**　本文提出基于局部近邻相关性的多标记算法，其从训练样例类标记与近邻样例中不同类别样例间欧式距离和个数两方面的相关性考虑，建立训练样例类标记与近邻样例中不同类别样例的欧式距离和个数间的映射关系。把训练样例与近邻间欧式距离平方的倒数作为近邻样例影响测试样例分类的权重，与测试样例距离小的样例影响测试样例分类的权值大，与测试样例距离大的样例影响测试样例分类的权值小。实验结果表明，ML-WKNN 处理多标记学习问题明显优于其它 4 种算法。今后还可从样例聚类和距离测度方法等方面对算法进行改进。

## 参 考 文 献


[1] McCallum Andrew. Multi-label text classification with a mixture model trained by EM[C]// AAAI'99 Workshop on Text Learning. 1999:1-7

[2] Schapire, Robert E, Singer Y. BoosTexter: A boosting-based system for text categorization[J]. Machine learning, 2000, 39(2/3):135-168

[3] Tsoumakas, Grigorios, Katakis I. Multi-label classification: An overview[J]. International Journal of Data Warehousing and Mining (IJDWM), 2007, 3(3):1-13

[4] Elisseeff, André, Weston J. A kernel method for multi-labelled classification[J]. Advances in neural information processing systems, 2001, 14:681-687

[5] Zhang Min-ling, Zhou Zhi-hua. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7):2038-2048

[6] 苗夺谦，卫志华. 中文文本信息处理的原理与应用[M]. 北京：清华大学出版社，2007:219-228

[7] Clare, Amanda, Ding K R. Knowledge discovery in multi-label phenotype data[C]// Principles of Data Mining and Knowledge Discovery. 2001:42-53

[8] Comité D, Francesco, Gilleron R, et al. Learning multi-label alternating decision trees from texts and data[C]// Machine Learning and Data Mining in Pattern Recognition. 2003:35-49

[9] Boutell, Matthew, Luo Jie-bo, et al. Learning multi-label scene classification[J]. Pattern recognition, 2004, 37(9):1757-1771

[10] Zhang Min-ling, Zhou Zhi-hua. Multilabel neural networks with applications to functional genomics and text categorization[J]. Knowledge and Data Engineering, 2006, 18(10):1338-1351

[11] Mitchell T M. Machine Learning [M]. USA: The McGrawHill Companies, Inc. 1997:165-177

[12] 张学工. 模式识别[M]. 北京：清华大学出版社，2010:120-130

[13] Boutell, Matthew, Luo Jie-bo, et al. Learning multi-label scene classification[J]. Pattern recognition, 2004, 37(9):1757-1771