

doi:10.11835/j.issn.1000-582X.2019.03.006

基于邻域维护准则的特征选择算法优化研究

刘 云,肖 雪

(昆明理工大学 信息工程与自动化学院,昆明 650500)

摘要:应用特征选择处理多标签数据分类时“维度灾难”问题已成为重要研究方向,因此提出一种基于邻域维护准则的特征选择算法(NPFS, feature selection algorithm based on neighborhood preservation criterion)。通过近似基于特征子空间和基于标签空间的 2 个相似度矩阵来构建相似性维护表达式,再通过线性近似扩展相似性维护公式得到邻域关系维护公式,并计算出邻域关系维护得分(NRPS, neighborhood relationship preserving score)来评估特征子集的重要性,结合贪婪方法设计具有 NRPS 的多标签特征选择算法(NPFS)。仿真结果表明,对比 MMIFS 算法和 MDMR 算法,所提出的算法在平均准确率、覆盖率、汉明损失、1-错误率、排名损失 5 个性能指标上均有改善。

关键词:特征选择;多标签分类;邻域关系维护得分;贪婪算法

中图分类号:TP311

文献标志码:A

文章编号:1000-582X(2019)03-058-07

Research on optimization of feature selection algorithm based on neighborhood preservation criterion

LIU Yun, XIAO Xue

(School of Information Engineering and Automation, Kunming University of
Science and Technology, Kunming 650500, P.R.China)

Abstract: The application of feature selection to deal with “Dimensional Disaster” in multi-label data classification has become an important research direction, we proposed a feature selection algorithm based on neighborhood preservation criterion (NPFS). A similarity preservation expression was constructed by approximating two similarity matrices based on feature subspace and label space. Then, the similarity preservation formulation was extended by the linear approximation to obtain a formulation of the neighborhood relationship preservation, and the importance of the feature subset was evaluated by calculating the neighborhood relationship preserving score (NRPS). A multi-label feature selection algorithm with NRPS was designed in combination with the greedy method(NPFS). The simulation results show that the metrics of average precision, coverage, hamming loss, one-error, ranking loss obtained by the proposed algorithm have been improved compared with those obtained by MMIFS algorithm and MDMR algorithm.

Keywords: feature selection; multi-label classification; neighborhood relationship preserving score; greedy algorithm

收稿日期:2018-10-11

基金项目:国家自然科学基金资助项目(61262040)。

Supported by National Natural Science Foundation of China(61262040).

作者简介:刘云(1973—)男,副教授,主要从事无线通信研究,(E-mail)liuyun@kmust.edu.cn。

随着海量商业数据的出现,在很多实际应用中的数据实例通常与多个标签关联^[1-2]。如何从多标签商业数据中通过特征选择来降低数据维度成为一项重要而具有挑战性的研究任务^[3-4]。针对此类问题的研究取得许多成果,传统算法通过特征和标签之间的相关性准则来选择有效特征子集^[5-6]。

Lee 等^[7]提出了一种基于多元互信息的多标签特征选择算法(MMIFS, feature selection for multi-label classification using multivariate mutual information),通过将高维熵的计算分解为多元互信息的累积和,最大化所选特征和标签之间的相关性来选择有效的特征子集。该算法只考虑特征和标签之间的相关性对特征选择的影响,忽略特征之间冗余性对算法性能的影响。Lin 等^[8]提出了基于最大依赖性和最小冗余性的多标签特征选择算法(MDMR, multi-label feature selection based on max-dependency and min-redundancy),通过结合最大化特征与标签之间的依赖性,以及最小化候选特征与选择特征之间的冗余性两个准则来进行特征选择,虽然这种策略能更好地提取出重要特征,但是没有考虑特征子空间中样本和标签空间中样本的邻域关系对算法性能的影响,从而影响多标签分类性能。

为了更有效地进行特征选择,文中提出基于邻域维护准则(称为领域关系维护得分 NRPS, neighborhood relationship preserving score)的特征选择算法(NPFS, feature selection algorithm based on neighborhood preservation criterion)。首先,利用基于特征子空间的样本相似度矩阵近似基于标签空间样本相似度矩阵来构造相似性维护表达式;其次,通过线性近似,样本相似度的顺序大致保持不变,则具有相似标签空间邻域关系的样本具有相似的特征子空间邻域关系,得到邻域关系维护公式,并计算得出邻域关系维护得分(NRPS)来评估特征子集的重要性;最后,结合贪婪算法^[9],设计具有 NRPS 的多标签特征选择算法(NPFS)。实验结果表明,对比 MMIFS 算法和 MDMR 算法, NPFS 算法在多标签分类的性能指标上均有提升。

1 基于邻域维护准则的多标签特征评估准则

1.1 相似性维护准则

文献[10]中有许多通过保留样本相似性的特征评估标准,例如拉普拉斯分数、Fisher 分数和 ReliefF。这些标准可以统一成

$$SC(f) = \hat{f}^T \hat{\mathbf{K}} \hat{f}, \quad (1)$$

式中: $\hat{\mathbf{K}}$ 是相似度矩阵; \hat{f} 是归一化的特征向量。对相似矩阵 \mathbf{K} 进行定义为

$$\min_{\hat{\mathbf{X}}} \|\mathbf{K}' - \mathbf{K}\|_F^2, \quad (2)$$

$$\mathbf{K}' = \hat{\mathbf{X}}^T \hat{\mathbf{X}},$$

式中, $\hat{\mathbf{X}}$ 表示所有特征中的所选 k 个特征,而 $\|\cdot\|_F$ 表示范数。相似度矩阵 \mathbf{K} 可以通过无监督学习中的某个距离度量或监督学习中的类别标签信息来构造。矩阵 $\mathbf{M} \in \mathbb{R}^{m \times n}$ 的范数定义为

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m M_{ij}^2}. \quad (3)$$

1.2 邻域维护准则

为了量化海量多标签商业数据,用 $\mathbf{X} \in \mathbb{R}^d$ 表示 d 维特征空间, $C = \{c_1, \dots, c_m\}$ 表示标签集, $T = \{x_i, Y_i \mid 1 \leq i \leq n\}$ 表示具有 n 个实例的多标签训练集,其中 $x_i \in \mathbf{X}$ 是一个 d 维特征向量 $[x_{i1}, \dots, x_{id}]$ 和 $Y_i \subseteq C$ 是与 x_i 相关的标签集合。子集 Y_i 可描述为一个 m 维二进制向量 $\mathbf{y}_i = [y_{i1}, \dots, y_{im}]$,其中 $y_{ij} = 1$ (Y_i 有标签 c_j), 否则为 0^[8]。

特征选择是选择一个特征子集近似表示所有特征。 $I \subseteq \{1, \dots, d\}$ 表示所选特征的索引集合,特征选择后的一个实例 $x \in \mathbb{R}^d$ 近似表示为 $x^I \in \mathbb{R}^{|I|}$,其中 x^I 是 x 的子向量。根据训练集 T 和索引集 I ,构造基于特征子空间和基于标签空间的 2 个样本相似度矩阵。对于 $\{x_1^I, \dots, x_n^I\}$,定义基于特征子空间的样本相似度矩阵为 $\mathbf{F}^I = [F_{ij}^I]_{n \times n}$, $F_{ij}^I = \langle x_i^I, x_j^I \rangle$ 。对于 $\{y_1, \dots, y_n\}$,定义基于标签空间的样本相似度矩阵为 $\mathbf{L} = [L_{ij}]_{n \times n}$, $L_{ij} = \langle y_i, y_j \rangle$ 。如果一对样本包含更多相同的标签,则认为它们更相似,根据文献[10]构建相似性维护公

式为

$$\min_I \| \mathbf{F}^I - \mathbf{L} \|_F^2. \quad (4)$$

假设子空间中每对样本的相似度近似等于标签空间中的相似度。然而,它们之间的这两种相似可能有不同的尺度。从另一个角度考虑,考虑每种空间中样本相似性的顺序,在每个样本相似度矩阵中,一对样本的相似度越大,则成对样本越接近。因此,样本相似度的顺序能很好地表达样本之间的邻近关系。利用基于特征子空间的样本相似度矩阵来近似基于标签空间样本相似度矩阵并构建相似性维护表达式。通过线性近似,样本相似性的顺序大致保持不变,则表明在标签空间中具有相似邻域关系的样本在特征子空间也具有相似的邻域关系。根据以上分析,构建邻域关系维护表达式为

$$\min_{I,a,b} \frac{1}{2} \| a\mathbf{F}^I + b\mathbf{1} - \mathbf{L} \|_F^2, \quad (5)$$

式中, a 和 b 是求解的 2 个变量, $\mathbf{1} = \mathbb{R}^{n \times n}$ 表示所有元素均为 1 的矩阵。为了从该公式获得特征评估标准,假设 I 已知,得:

$$\begin{aligned} \arg \min_{a,b} \frac{1}{2} \| a\mathbf{F}^I + b\mathbf{1} - \mathbf{L} \|_F^2 &= \arg \min_{a,b} \frac{1}{2} \text{tr}[(a\mathbf{F}^I + b\mathbf{1} - \mathbf{L})^\top (a\mathbf{F}^I + b\mathbf{1} - \mathbf{L})] = \\ &= \arg \min_{a,b} \frac{1}{2} \text{tr}(a^2 \mathbf{F}^I \mathbf{F}^I + 2ab \mathbf{F}^I \mathbf{1} + b^2 \mathbf{1}^2 - 2a \mathbf{F}^I \mathbf{L} - 2b \mathbf{1} \mathbf{L} + \mathbf{L}^2), \end{aligned} \quad (6)$$

令 $Z = \frac{1}{2} \| a\mathbf{F}^I + b\mathbf{1} - \mathbf{L} \|_F^2$, 分别求导计算 a, b 。

$$\frac{dZ}{da} = a \text{tr}(\mathbf{F}^I \mathbf{F}^I) + b \text{tr}(\mathbf{F}^I \mathbf{1}) - \text{tr}(\mathbf{F}^I \mathbf{L}) = 0, \quad (7)$$

$$\frac{dZ}{db} = a \text{tr}(\mathbf{F}^I \mathbf{1}) + b \text{tr}(\mathbf{1}^2) - \text{tr}(\mathbf{L} \mathbf{1}) = 0, \quad (8)$$

根据式(7)和式(8),推出

$$\begin{pmatrix} \text{tr}(\mathbf{F}^I \mathbf{F}^I) & \text{tr}(\mathbf{F}^I \mathbf{1}) \\ \text{tr}(\mathbf{F}^I \mathbf{1}) & \text{tr}(\mathbf{1}^2) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \text{tr}(\mathbf{F}^I \mathbf{L}) \\ \text{tr}(\mathbf{L} \mathbf{1}) \end{pmatrix}. \quad (9)$$

求解式(9),引入引理 1 和推论 1。

引理 1:柯西不等式^[11]:假设 $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \in \mathbb{R}$, 然后,

$$\left[\sum_{i=1}^n \alpha_i \beta_i \right]^2 \leq \sum_{i=1}^n \alpha_i^2 \sum_{i=1}^n \beta_i^2. \quad (10)$$

当且仅当 $\beta_i = 0$ 或 $\alpha_i = c\beta_i (i=1, \dots, n; c \in \mathbb{R})$ 时,等号成立。

推论 1:如果 \mathbf{A} 和 \mathbf{B} 都是对称矩阵,那么

$$(\text{tr}(\mathbf{AB}))^2 \leq \text{tr}(\mathbf{A}^2) \text{tr}(\mathbf{B}^2), \quad (11)$$

当且仅当 $\mathbf{B} = \mathbf{0}$ 或 $\mathbf{A} = c\mathbf{B} (c \in \mathbb{R})$ 时,等号成立。

显然,样本间的相似性在实际应用中并不完全相同。因此, $\mathbf{F}^I \neq c\mathbf{1} (c \in \mathbb{R})$, 那么

$$(\text{tr}(\mathbf{F}^I \mathbf{1}))^2 < \text{tr}(\mathbf{F}^I \mathbf{F}^I) \text{tr}(\mathbf{1}^2), \quad (12)$$

所以

$$\begin{vmatrix} \text{tr}(\mathbf{F}^I \mathbf{F}^I) & \text{tr}(\mathbf{F}^I \mathbf{1}) \\ \text{tr}(\mathbf{F}^I \mathbf{1}) & \text{tr}(\mathbf{1}^2) \end{vmatrix} \neq 0. \quad (13)$$

根据式(9)和式(13)

$$a = \frac{\text{tr}(\mathbf{F}^I \mathbf{L}) \text{tr}(\mathbf{1} \mathbf{1}) - \text{tr}(\mathbf{F}^I \mathbf{1}) \text{tr}(\mathbf{L} \mathbf{1})}{\text{tr}(\mathbf{F}^I \mathbf{F}^I) \text{tr}(\mathbf{1} \mathbf{1}) - \text{tr}(\mathbf{F}^I \mathbf{1}) \text{tr}(\mathbf{F}^I \mathbf{1})}, \quad (14)$$

$$b = \frac{\text{tr}(\mathbf{F}^I \mathbf{F}^I) \text{tr}(\mathbf{L} \mathbf{1}) - \text{tr}(\mathbf{F}^I \mathbf{1}) \text{tr}(\mathbf{F}^I \mathbf{L})}{\text{tr}(\mathbf{F}^I \mathbf{F}^I) \text{tr}(\mathbf{1} \mathbf{1}) - \text{tr}(\mathbf{F}^I \mathbf{1}) \text{tr}(\mathbf{F}^I \mathbf{1})}, \quad (15)$$

然后,得到一个称为邻域关系维护得分(NRPS)的特征评估标准,评估每个特征子集的重要性。对于每个索引集合 I , 计算索引集 I 对应的特征子集的 NRPS 为

$$\text{NRPS}(I) = \|\alpha \mathbf{F}^I + \mathbf{b}\mathbf{1} - \mathbf{L}\|_F^2, \quad (16)$$

该值越小,特征子集越重要。

2 NPFS 算法

令 $\text{SC}(r)$ 为第 r 个特征的邻域关系维护得分。对于第 r 个特征 ($r=1, \dots, d$, 其中 d 是特征的数量), 其分数值 $\text{SC}(r) = \text{NRPS}(\{r\})$ 。应用 NRPS 度量每个特征子集的重要性, 使用传统的前向搜索策略来选择特征, 设计一个用于特征选择的贪婪算法。首先, 将选择的特征索引集合 I 初始化为 \emptyset , 特征索引集合 D 初始化为 $\{1, \dots, d\}$ 。其次, 基于以下函数将 D 的元素逐一放入 I 中, 有

$$\arg \min_{r \in D} \text{NRPS}(I \cup \{r\}). \quad (17)$$

要选择 k 个特征, NPFS 算法需要运行 k 次迭代, 算法的计算复杂度为 $O(kdn^2)$ 。

NPFS 算法代码如下:

输入: 数据矩阵 $\mathbf{X} \in R^{n \times d}$, 多标签 $\mathbf{Y} \in R^{n \times m}$, \mathbf{K} (被选择的特征数目)。

输出: 所选特征的索引集合 $I \subseteq \{1, \dots, d\}$, 且 $|I| = K$ 。

- 1: 计算基于标签空间的相似度矩阵 $\mathbf{L}, \mathbf{L} = \mathbf{Y}\mathbf{Y}^T$;
- 2: $I = \emptyset, D = \{1, \dots, d\}, \mathbf{F}^I = \mathbf{0}$; // 初始化 I, D, \mathbf{F}^I
- 3: $\mathbf{X} = \mathbf{X}\mathbf{G}^{-1}$, 其中 $\mathbf{G} = \sqrt{\text{diag}(\mathbf{X}^T \mathbf{X})}$; // 标准化 \mathbf{X}
- 4: $\mathbf{C}_n = \mathbf{I}_n - \mathbf{1}\mathbf{1}^T/n$; // 计算中心矩阵
- 5: $\mathbf{X} = \mathbf{C}_n \mathbf{X}$; // 零均值化
- 6: 重复
- 7: 如果 $r \in D$
- 8: 计算 \mathbf{F}^r , 其中 $\mathbf{F}^r = \mathbf{F}^I + \mathbf{F}^{(r)}$;
- 9: $a = \frac{\text{tr}(\mathbf{F}^r \mathbf{L}) \text{tr}(\mathbf{1}\mathbf{1}) - \text{tr}(\mathbf{F}^r \mathbf{1}) \text{tr}(\mathbf{L}\mathbf{1})}{\text{tr}(\mathbf{F}^r \mathbf{F}^r) \text{tr}(\mathbf{1}\mathbf{1}) - \text{tr}(\mathbf{F}^r \mathbf{1}) \text{tr}(\mathbf{F}^r \mathbf{1})}$;
- 10: $b = \frac{\text{tr}(\mathbf{F}^r \mathbf{F}^r) \text{tr}(\mathbf{L}\mathbf{1}) - \text{tr}(\mathbf{F}^r \mathbf{1}) \text{tr}(\mathbf{F}^r \mathbf{L})}{\text{tr}(\mathbf{F}^r \mathbf{F}^r) \text{tr}(\mathbf{1}\mathbf{1}) - \text{tr}(\mathbf{F}^r \mathbf{1}) \text{tr}(\mathbf{F}^r \mathbf{1})}$;
- 11: $\text{SC}(r) = \|\alpha \mathbf{F}^r + b\mathbf{1} - \mathbf{L}\|_F^2$;
- 12: 结束;
- 13: $\mathbf{I} \leftarrow \mathbf{I} \cup \{r^*\}, \mathbf{D} \leftarrow \mathbf{D} - \{r^*\}, \mathbf{F}^I \leftarrow \mathbf{F}^{r^*}$, 其中 $r^* = \arg \min_r \text{SC}(r)$;
- 14: 直到 $|I| = K$;
- 15: 返回特征索引集 I 。

整个算法的具体过程如下:

Step1: 构建样本相似矩阵

$$\mathbf{F}^I = \mathbf{X}^I (\mathbf{X}^I)^T, \mathbf{L} = \mathbf{Y}\mathbf{Y}^T,$$

Step2: 构建邻域维护公式

$$\arg \min_{I, \alpha, b} \frac{1}{2} \|\alpha \mathbf{F}^I + b\mathbf{1} - \mathbf{L}\|_F^2,$$

Step3: 获取特征评估标准

$$\text{NRPS}(I) = \min_{\alpha, b} \frac{1}{2} \|\alpha \mathbf{F}^I + b\mathbf{1} - \mathbf{L}\|_F^2,$$

Step4: 设计特征选择算法。

3 仿真分析

3.1 数据集和评价指标

为了验证所提出的算法, 文中选取来自 Mulan Library^[12-13] 商业领域的多标签数据集进行实验, 给定一

个多标签数据集 $T = \{(x_i, Y_i) | 1 \leq i \leq p\}$, 用 $\dim(T)$, $L(T)$, $|T|$ 分别表示特征数量, 总标签数量和实例数量。表 1 为仿真数据集详细信息。

表 1 仿真数据集
Table 1 The simulation dataset

数据集	$\dim(T)$	$L(T)$	训练集 $ T $	测试集 $ T $
Business	438	30	2 000	3 000

多标签学习系统的性能评估不同于单标签系统, 多标签学习系统评估指标更复杂。对指标进行预定义, 一个测试数据集 $S = \{(x_i, Y_i) | 1 \leq i \leq p\}$, 给定一个实例 x_i , 由多标签分类器预测的标签集合表示为 $h(x_i)$, 标签 y 预测的排序表示为 $\text{rank}_i(y)$ 。文中选用 5 个常用多标签学习算法评价指标。

1) 平均准确率 (AP, average precision)

$$\text{AP}(S) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{\text{Prec}_i(y)}{\text{rank}_i(y)}, \quad (18)$$

其中

$$\text{Prec}_i(y) = |\{y' | \text{rank}_i(y') \leq \text{rank}_i(y), y' \in Y_i\}|,$$

该指标评估平均有多少实际标签比给定测试样本包含的标签高。值越大则算法越优。

2) 覆盖率 (CV, coverage)

$$\text{CV}(S) = \frac{1}{p} \sum_{i=1}^p \text{argmaxrank}_i(y) - 1, \quad (19)$$

该指标评估要囊括测试样本实际包含的标签平均需要的最大排序距离。值越小则算法越优。

3) 汉明损失 (HL, Hamming loss)

$$\text{HL}(S) = \frac{1}{p} \sum_{i=1}^p \frac{1}{m} |h(x_i) \otimes Y_i|, \quad (20)$$

其中 \otimes 表示 2 个集合的对称差异, 该指标评估有多少实例标签对被错误分类。也就是说, 属于该实例的标签不被预测或者不属于该实例的标签被预测。值越小则算法越优。

4) 1-错误率 (OE, one-error)

$$\text{OE}(S) = \frac{1}{p} \sum_{i=1}^p \delta(\text{argminrank}_i(y) \notin Y_i), \quad (21)$$

该指标评估排名最高的标签不在相关标签集中的次数。值越小则算法越优。

5) 排名损失 (RL, ranking loss)

$$\text{RL}(S) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|Y_i| |\bar{Y}_i|} \times |\{(y, y') \in Y_i \times \bar{Y}_i | \text{rank}_i(y) > \text{rank}_i(y')\}|, \quad (22)$$

式中 \bar{Y}_i 表示 Y_i 相对于 C 的互补集合。该指标评估有多少不相关标签排名高于相关标签的标签对。值越小则算法越优。

3.2 算法对比仿真分析

将所提出的 NPFS 算法与 MMIFS 算法和 MDMR 算法进行实验, 并选择分类算法 ML-KNN^[14] 作为评估的分类器, 将最近邻 K 的数量和平滑值 S 分别被设置为 10 和 1。使用所有特征的 5%~40% 范围内的选定特征进行实验, 间隔为 5%。图 1~图 5 分别给出 3 种算法在 Business 数据集的 5 项评价指标随着特征数 K 增加的变化曲线。

如图 1 所示, 随着所选特征数目增加, NPFS 算法的平均准确率最高, MDMR 算法平均准确率次之, MMIFS 算法平均准确率最低。

如图 2 所示, 随着所选特征数目增加, NPFS 算法覆盖率最小, MDMR 算法覆盖率次之, MMIFS 算法覆盖率最高。

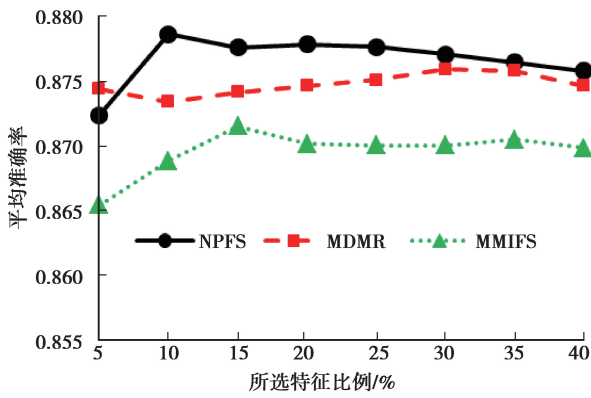


图 1 不同算法平均准确率对比图

Fig. 1 Comparing average precision in different algorithms

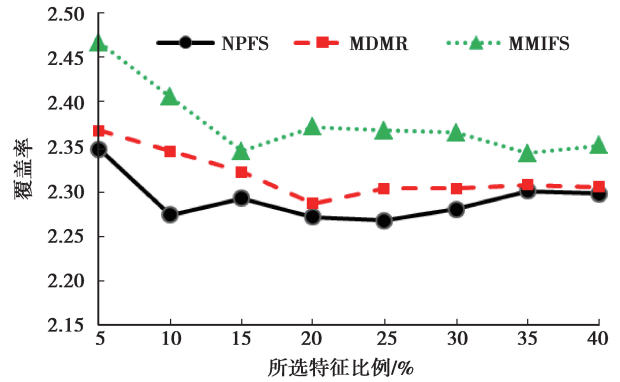


图 2 不同算法覆盖率对比图

Fig. 2 Comparing coverage in different algorithms

如图 3 所示,随着所选特征数目增加,NPFS 算法的汉明损失最小,MDMR 算法平均准确率次之,MMIFS 算法汉明损失最大。

如图 4 所示,随着所选特征数目增加,NPFS 算法的 1-错误率最小,MDMR 算法 1-错误率次之,MMIFS 算法 1-错误率最大,特征比例增加到 30%后,虽然错误率有所增加,但相比于另外 2 种算法性能也是最好的。

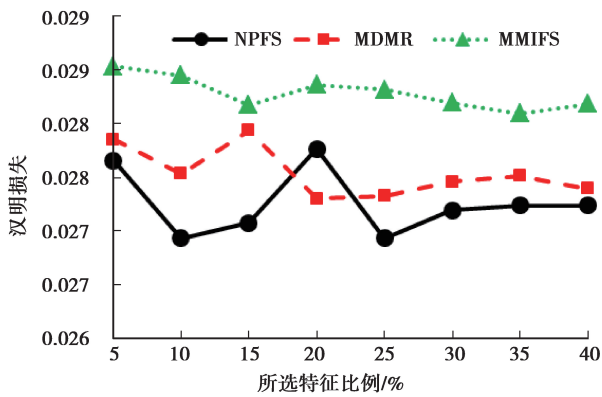


图 3 不同算法汉明损失对比图

Fig. 3 Comparing Hamming loss in different algorithms

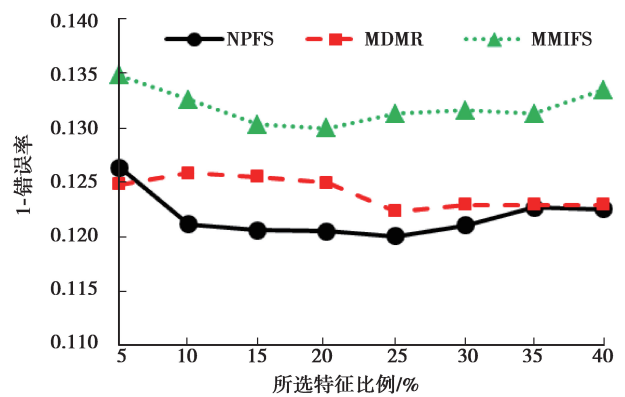


图 4 不同算法 1-错误率对比图

Fig. 4 Comparing one-error in different algorithms

如图 5 所示,随着所选特征数目增加,NPFS 算法的排名损失最小,MDMR 算法排名损失次之,MMIFS 算法排名损失最大。

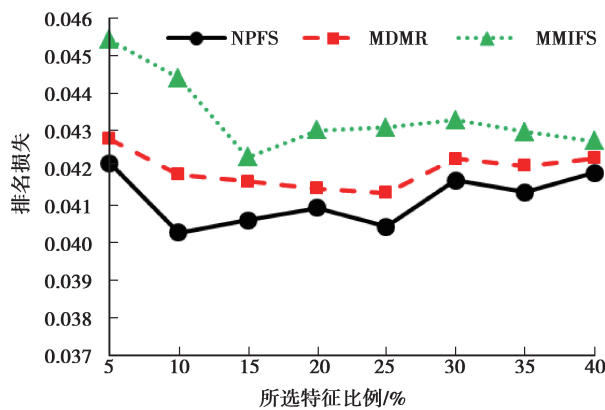


图 5 不同算法排名损失对比图

Fig. 5 Comparing ranking loss in different algorithms

4 结 论

为了从海量商业多标签数据中有效地选择重要特征来降低数据维度,提出一种基于邻域维护准则的特征选择算法(NPFS)。首先,通过近似基于特征子空间和基于标签空间的 2 个相似度矩阵来构建相似性维护表达式;其次,通过线性近似扩展相似性维护公式得到邻域关系维护公式,并计算得出邻域关系维护得分(NRPS)来评估特征子集的重要性;最后,结合贪婪方法设计具有 NRPS 的多标签特征选择算法。实验结果表明,对比 MMIFS 算法和 MDMR 算法,在标准商业数据集中,NPFS 算法在平均准确率、覆盖率、汉明损失、1-错误率、排名损失 5 个性能指标上均有提升。文中使用简单的线性关系构造样本相似度矩阵,下一步工作将尝试扩展一般非线性关系构造样本相似度矩阵。

参考文献:

- [1] Zhang M L, Wu L. Lift: multi-label learning with label-specific features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107-120.
- [2] Zhang M L, Zhou Z H. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [3] Huang J, Li G R, Huang Q M, et al. Learning label-specific features and class-dependent labels for multi-label classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309-3323.
- [4] Glinka K, Woźniak R, Zakrzewska D. Improving multi-label medical text classification by feature selection [C] //2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Poznan, 2017: 176-181.
- [5] He Z F, Yang M, Liu H D. Multi-task joint feature selection for multi-label classification [J]. Chinese Journal of Electronics, 2015, 24(2): 281-287.
- [6] Xu C, Liu T L, Tao D C, et al. Local rademacher complexity for multi-label learning [J]. IEEE Transactions on Image Processing, 2016, 25(3): 1495-1507.
- [7] Lee J, Kim D W. Feature selection for multi-label classification using multivariate mutual information [J]. Pattern Recognition Letters, 2013, 34(3): 349-357.
- [8] Lin Y J, Hu Q H, Liu J H, et al. Multi-label feature selection based on max-dependency and min-redundancy [J]. Neuro computing, 2015, 168: 92-103.
- [9] Pathania A, Venkatramani V, Shafique M, et al. Optimal greedy algorithm for many-core scheduling [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2017, 36(6): 1054-1058.
- [10] Zhao Z, Wang L, Liu H, et al. On similarity preserving feature selection [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(3): 619-632.
- [11] Mutuk H. A physical proof for cauchy's inequality[J]. The American Mathematical Monthly, 2016, 123(8): 843.
- [12] 李峰, 苗夺谦, 张志飞, 等. 基于互信息的粒化特征加权多标签学习 k 近邻算法[J]. 计算机研究与发展, 2017, 54(5): 1024-1035.
LI Feng, MIAO Duoqian, ZHANG Zhifei, et al. Mutual information based granular feature weighted k-nearest neighbors algorithm for multi-label learning[J]. Journal of Computer Research and Development, 2017, 54(5): 1024-1035. (in Chinese)
- [13] Xu H T, Xu L Y. Multi-label feature selection algorithm based on label pairwise ranking comparison transformation[C]//2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017: 1210-1217.
- [14] Zhang M L, Zhou Z H. ML-KNN: a lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048.