Computer Science and Application 计算机科学与应用, 2024, 14(10), 10-21
Published Online October 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.1410198

Hans汉斯

基于三元互信息的成对多标签特征选择算法研究

张　平[1,2,3]，王光磊[2,3]，张亚娟[2,3*]，曹　宇[2,3]

[1]河北工业大学省部共建电工装备可靠性与智能化国家重点实验室，天津
[2]河北工业大学人工智能与数据科学学院，天津
[3]河北省大数据计算重点实验室，天津

收稿日期：2024年8月25日；录用日期：2024年9月23日；发布日期：2024年9月30日

摘　要

基于信息论的特征选择算法在度量候选特征所提供的分类信息时，往往仅考虑单一标签的情况，忽略了候选特征和成对标签存在的多样关联关系，这可能导致低估了候选特征的重要性。为解决这一问题，提出一种新颖的基于三元互信息的成对多标签特征选择算法(Pairwise multi-label feature selection based on interaction mutual information, IPFS)。具体地，IPFS算法为不同的成对标签分配基于三元互信息的不同权重，并据此权重测量候选特征为两个标签提供的分类信息总量，从而精确评估候选特征的重要性，同时基于最大相关最小冗余原则，筛选出最优的特征子集。最后，将提出的算法与其他8个先进的特征选择算法在12个多样化的数据集上进行了比较。实验结果表明，IPFS在3个评估指标上均显著优于其他算法。

关键词

机器学习，特征选择，三元互信息，分类

Pairwise Multi-Label Feature Selection Method Based on Interaction Mutual Information

Ping Zhang[1,2,3], Guanglei Wang[2,3], Yajuan Zhang[2,3*], Yu Cao[2,3]

[1]State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin
[2]School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin
[3]Hebei Province Key Laboratory of Big Data Calculation, Tianjin

[*]通讯作者。

文章引用: 张平, 王光磊, 张亚娟, 曹宇. 基于三元互信息的成对多标签特征选择算法研究[J]. 计算机科学与应用, 2024, 14(10): 10-21. DOI: 10.12677/csa.2024.1410198

Received: Aug. 25th, 2024; accepted: Sep. 23rd, 2024; published: Sep. 30th, 2024

Abstract

The feature selection methods based on information theory usually focus on considering the single label when evaluating the classification information provided by the candidate features, and do not take into account the multiple correlations between the candidate features and the paired labels, thus underestimating the importance of the candidate features. To solve this issue, an innovative paired multi-label feature selection method based on interaction mutual information (IPFS) was proposed. Specifically, IPFS method assigns different weights based on interaction mutual information to different pairs of labels, so as to accurately evaluate the importance of candidate features, and further select the most suitable feature subset based on the maximum correlation minimum redundancy strategy. To verify the effectiveness of the proposed method, IPFS is compared with eight other advanced feature selection methods on 12 diverse datasets, and the results show that IPFS significantly outperforms other methods on four different evaluation metrics.

Keywords

Machine Learning, Feature Selection, Interaction Mutual Information, Classification

1. 引言

在当今信息时代，科技的飞速发展带来了海量的高维多标签数据，这些数据在多个领域[1]中不断涌现，如文本分析、图像识别和生物信息学等。然而，高维数据的存储和计算挑战日益凸显，如何有效地处理这些数据成为了一个迫切需要解决的问题。和其他方法不同的是，特征选择[2]-[4]在删除无关和冗余特征时，可以最大程度保留原始物理特征，同时实现降维效果。这样可以有效提高数据质量以及分类性能，使得模型训练时间大幅减少的同时提升模型的可解释性。

常见的基于信息论特征选择方法策略遵循最大相关最小冗余策略，即所选特征子集和标签集合具有最大相关性和最小冗余性。在此基础上，衍生出很多优秀特征评估方法，如 PMU [5]、D2F [6]、SCLS [7]、MIFS [8]、FIMF [9]、LRFS [10]和 FSSL [11]等。这些方法通过去除冗余特征来提高特征子集的整体质量。但是在评估相关性方面，这些方法都仅仅考虑了候选特征对单个标签的信息贡献，例如 D2F、PMU、SCLS、LRFS 等。然而，为了更精确地测量相关性，评估时还应考虑候选特征和成对标签之间的相关性。例如，在文本分析领域，特征词“温度”与标签“天气”具有相关性。若同时考虑另一标签“季节”，则“温度”与这两个标签的联合相关性将显著增强，进而提升其重要性。这种综合考虑两个标签的方法可以更全面地评估特征词的分类潜力，从而在特征选择过程中识别出更具信息量的特征。

在前文基础上，本文提出了一种基于三元互信息的成对多标签特征选择算法(Pairwise multi-label feature selection based on interaction mutual information, IPFS)。主要贡献包括：1) 标签对和不同的候选特征，利用三元互信息赋予不同的权重；2) 通过该权重准确测量候选特征为成对标签提供的分类信息，从而得到候选特征的重要性；3) 与 8 个特征选择算法在 12 个不同的数据集上进行比较。实验结果表明，提出

算法在分类性能上更具优势。

2. 相关理论

在本章中，我们主要介绍一下信息论[12]相关知识。信息熵是用来评估随机变量不确定性的有效度量。假设有随机变量 $X=\{x_1,x_2,x_3,\cdots,x_n\}$， $p(x_i)$ 是 x_i 的概率密度，则信息熵 $H(X)$ 的定义如下：

$$H(X)=-\sum_{i=1}^{n}p(x_i)\log p(x_i) \tag{1}$$

这里的 log 底数为 2。互信息用于度量两个随机变量共享的信息量。对于两个随机变量，互信息的含义是指一个随机变量由于已知另一个随机变量而减少的信息量大小假设有另一随机变量 $Y=\{y_1,y_2,y_3,\cdots,y_n\}$， $p(x_i,y_j)$ 为 x_i 和 y_j 的联合概率密度，则互信息的定义如下：

$$I(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{n}p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)p(y_j)} \tag{2}$$

联合互信息是一种用于度量两个随机变量联合起来和另一个随机变量之间相互依赖程度的指标。它是基于互信息定义的一种推广，用于衡量多变量之间的关联性。联合互信息与熵的关系如下：

$$I(X,Y;Z)=H(Z)-H(Z\mid X,Y)=H(X,Y)-H(X,Y\mid Z) \tag{3}$$

三元互信息 $I(X;Y;Z)$ 计算三个随机变量共享的信息量。其公式定义如下所示：

$$\begin{aligned}I(X;Y;Z)&=I(X;Y)-I(X;Y\mid Z)\\&=I(X;Z)-I(X;Z\mid Y)\\&=I(Y;Z)-I(Y;Z\mid X)\\&=I(X;Z)+I(Y;Z)-I(X,Y;Z)\end{aligned} \tag{4}$$

相对于联合互信息衡量两个变量和另一变量之间的依赖程度，三元互信息可以衡量三个随机变量之间的相互依赖程度，三元互信息越大，代表三个变量越紧密，共享的信息越多。

3. 算法分析

基于信息论的多标签特征选择算法普遍采用“最大相关性–最小冗余性”策略，即选择的特征集合与标签集合保持最大相关性，同时尽量减小特征之间的冗余性。这一策略可以通过以下目标函数公式来概括：

$$J(f_k)=rel(f_k;L)-\beta*ref(f_k;S) \tag{5}$$

上式中， $J(f_k)$ 表示评估函数， f_k 是候选特征， L 是标签集合， $rel(f_k;L)$ 表示候选特征和标签集合之间的相关性， S 是已选特征集合， $ref(f_k;S)$ 表示候选特征和已选特征之间的冗余性， β 用于平衡相关性和冗余性之间权重的超参数。通过前向搜索策略，选择使 $J(f_k)$ 取得最大值的特征，并将其加入到已选特征集合中，直到满足终止条件。

现有的多标签特征选择算法在计算候选特征与标签集合的相关性时，普遍采用候选特征 f_k 和多标签集合 L 中每个标签的互信息累加和($\sum_{l_i\in L}I(f_k;l_i)$)来计算，例如 D2F、PMU、SCLS 等。这些算法主要衡量每个候选特征与单个标签之间的相关性，忽略了候选特征向成对标签提供的分类信息，进而可能导致对候选特征重要性的评估不够准确。进一步分析，在考虑候选特征为成对标签提供的分类信息时，也应该根据不同关系的成对标签赋予不同的权重。

3.1. 特征相关性度量分析

我们通过图 1 分析在多标签特征选择过程中，特征和成对标签之间相关性的分析。

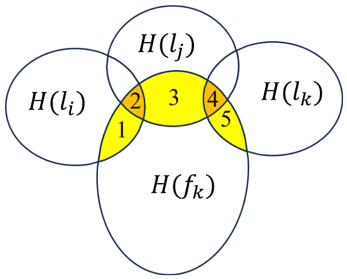


Figure 1. The relationship between feature f_k and different paired labels

图 1. 特征 f_k 与不同标签对之间的关系

在图 1 中，f_k 是候选特征，l_i，l_j 和 l_k 是三个标签。区域 1 + 2 + 3 + 4 是候选特征 f_k 为标签(l_i，l_j)提供的信息量，即联合互信息 $I(l_i,l_j;f_k)$，区域 2 + 3 + 4 + 5 是候选特征 f_k 为标签(l_j，l_k)提供的信息量，即联合互信息 $I(l_j,l_k;f_k)$。考虑当区域 1 等于区域 5 时，即满足 $I(l_i,l_j;f_k)=I(l_j,l_k;f_k)$，这意味着候选特征为($l_i$，$l_j$)和($l_j$，$l_k$)提供的信息量相同。然而，由图中可知，区域 2 和区域 4 分别是标签对(l_i，l_j)和(l_j，l_k)冗余的部分，并且区域 2 小于区域 4，即 $I(f_k;l_i;l_j)<I(f_k;l_j;l_k)$，这部分不能被候选特征提供的信息加以区分。实际上，候选特征 f_k 为标签(l_i，l_j)提供的有效信息要大于给标签(l_j，l_k)提供的有效信息。

根据以上讨论，我们提出一个基于三元互信息的权重：

$$w_{i,j,k}=1-\frac{I(f_k;l_i;l_j)}{H(l_i)+H(l_i)} \tag{6}$$

其中，$0<w<2$，证明如下：

由信息论可知：

$$\begin{aligned} I(f_k;l_i;l_j) &= I(f_k;l_i)+I(f_k;l_j)-I(l_i,l_j;f_k) \\ &\le I(f_k;l_i)+I(f_k;l_j) \\ &= H(l_i)-H(l_i\mid f_k)+H(l_j)-H(l_j\mid f_k) \\ &\le H(l_i)+H(l_j) \end{aligned} \tag{7}$$

所以 $I(f_k;l_i;l_j)$ 满足：

$$I(f_k;l_i;l_j)\le H(l_i)+H(l_j) \tag{8}$$

又因为：

$$\begin{aligned} I(f_k;l_i;l_j) &= I(f_k;l_i)+I(f_k;l_j)-I(l_i,l_j;f_k) \\ &\ge -I(l_i,l_j;f_k) \\ &= -H(l_i,l_j)+H(l_i,l_j\mid f_k) \\ &\ge -H(l_i,l_j) \\ &\ge -H(l_i)-H(l_j) \end{aligned} \tag{9}$$

所以：

$$I\left(f_k;l_i;l_j\right)\geq -H\left(l_i\right)-H\left(l_j\right) \tag{10}$$

进而可以得到：

$$-H\left(l_i\right)-H\left(l_j\right)\leq I\left(f_k;l_i;l_j\right)\leq H\left(l_i\right)+H\left(l_j\right) \tag{11}$$

即

$$-1\leq \frac{I\left(f_k;l_i;l_j\right)}{H\left(l_i\right)+H\left(l_j\right)}\leq 1 \tag{12}$$

因此权重 $w_{i,j,k}$ 的取值范围是 $[0,2]$。当 $w=1$ 时，即 $I\left(f_k;l_i;l_j\right)=0$，说明候选特征为两个标签提供的信息量等于给两个标签单独提供的信息量。当 $w<1$ 时，即 $I\left(f_k;l_i;l_j\right)>0$，说明特征为两个标签提供的信息中有一部分是无效信息。当 $w>1$ 时，即 $I\left(f_k;l_i;l_j\right)<0$，由三元互信息定义可知特征 f_k 为两个标签提供的信息量要大于给两个标签单独提供的信息量。

基于以上分析，提出对特征相关性测量的新定义如下：

新特征相关性：设 F 是特征全集，S 是已选特征集合，L 是标签集合，$f_k\in F-S$ 表示候选特征，$l_i,l_j\in L$ 是两个标签。那么新特征相关性的定义如下：

$$NREL\left(f_k;L\right)=\sum_{l_i\in L}\left\{I\left(f_k;l_i\right)+\frac{1}{|L|}\sum_{l_j\in L,l_i\neq l_j}w*I\left(l_i,l_j;f_k\right)\right\} \tag{13}$$

$NREL\left(f_k;L\right)$ 在衡量候选特征相关性时，使用 $I\left(f_k;l_i\right)$ 计算候选特征 f_k 为单个标签提供的信息量，同时使用 $w*I\left(l_i,l_j;f_k\right)$ 来计算候选特征为两个标签提供的有效信息量。使用权重 $1/|L|$ 来平衡两种信息量。

综合上述分析，将候选特征为成对标签提供的信息量纳入候选特征相关性的评估机制中，能够更准确地评估其重要性。同时，在该过程中引入权重机制来有效挖掘候选特征为两个标签提供的有效信息。

3.2. 提出算法

利用上一节提出的新特征相关性以及最大相关最小冗余策略，原始的目标函数(5)可以写成

$$\begin{aligned}J\left(f_k\right)&=NREL\left(f_k;L\right)-\sum_{f_i\in S}I\left(f_k;f_i\right)\\&=\sum_{l_i\in L}\left\{I\left(f_k;l_i\right)+\frac{1}{|L|}\sum_{l_j\in L,l_i\neq l_j}w*I\left(l_i,l_j;f_k\right)\right\}-\sum_{f_i\in S}I\left(f_k;f_i\right)\end{aligned} \tag{14}$$

在该式中，S 是已选特征集合，f_i 表示集合 S 中的特征。第一项使用 $NREL\left(f_k;L\right)$ 来衡量特征相关性，第二项使用 $\sum_{f_i\in S}I\left(f_k;f_i\right)$ 作为特征冗余项。在特征选择过程中使用前向搜索策略，即每次迭代选择一个获得 $J\left(f_k\right)$ 最大值的特征加入到已选特征集合 S 中。算法伪代码如下：

输入：原始特征集合 F，标签集合 L，特征个数 K

输出：选择的特征对应的索引集合 S

1、$S\leftarrow\varnothing$

2、$k=0$

3、for each $f_k\in F$

续表

4、 计算 $NREL(f_k;L)$

5、end for

6、While $k<K$

7、 if $k=0$ then

8、 $f_j=\arg\max(NREL(f_k;L))$

9、 $S=S\cup\{f_j\}$

10、 $k=k+1$

11、 $F=F-\{f_j\}$

12、 End if

13、 For each candidate feature $f_i\in F$ do

14、 计算 $J(f_i)$

15、 End for

16、 $f_j=\text{argmax}(J(f_i))$

17、 $S=S\cup\{f_j\}$

18、 $k=k+1$

19、 $F=F-\{f_j\}$

20、End While

首先。初始化每个参数：已选特征集合 S 和特征个数 K，然后，3~12 行计算每个特征的 $NREL(f_k;L)$ 并选择最大值作为第一个特征加入到 S 中。最后，13~20 行计算公式(14)并选择满足要求的特征直到满足阈值 K。

4. 实验

4.1. 实验设置

为了验证所提出算法的有效性，本文将 IPFS 算法与八种算法(MIFS、D2F、PMU、SCLS、LRFS、FIMF、FSSL、AIII-FS [13])在 12 个公开数据集上进行实验比较。其中，MIFS、D2F、PMU、SCLS、LRFS、FIMF 是传统的多标签特征选择算法，FSSL 是流标签场景下的特征选择算法，AIII-FS 是最新提出的基于信息论的多标签特征选择算法。实验中的 12 个数据集来自 Mulan 公开数据库[14]，数据集的信息如表 1 所示。这些数据集包含三个不同的领域，其中 10 个多标签数据集(Medical、Enron、Arts、Business、Educations、Entertain、Recreation、Reference、Science 和 Social)被广泛运用于文本分类，图形数据集 Scene 对图像样本进行语义分类，Yeast 数据集被用于进行生物信息分类。

4.2. 实验结果分析

表 2~4 展示九种特征选择算法在 12 个不同数据集上的性能表现对比，评价指标分别是 Hamming Loss、Zero One Loss 和 Macro-F1。其中，Hamming Loss 和 Zero One Loss 指标是基于 ML-KNN [15]分类

Table 1. Description of data sets
表 1. 数据集描述

数据集	样本数	特征数	标签数	领域
Medical	978	1449	45	Text
Scene	2407	294	6	images
Yeast	2417	103	14	Biology
Enron	1702	1001	53	Text
Arts	5000	462	26	Text
Business	5000	438	30	Text
Educations	5000	550	33	Text
Entertain	5000	640	21	Text
Recreation	5000	606	22	Text
Reference	5000	793	33	Text
Science	5000	743	40	Text
Social	5000	1047	39	Text

Table 2. Comparison results of 9 multi-label feature selection algorithms on Hamming Loss index
表 2. 9 个多标签特征选择算法在 Hamming Loss 指标上的比较结果

数据集	IPFS	MIFS	D2F	PMU	SCLS	LRFS	FIMF	FSSL	AIII-FS
Medical	**0.0157 ± 0.001**	0.0165 ± 0.0021	0.0196 ± 0.001	0.0197 ± 0.0011	0.0233 ± 0.0002	0.0175 ± 0.001	0.0174 ± 0.001	0.0184 ± 0.0024	0.0218 ± 0.0001
Scene	0.1379 ± 0.0119	0.1704 ± 0.0097	0.1492 ± 0.0064	0.1473 ± 0.0066	0.1734 ± 0.003	0.1419 ± 0.0099	0.1663 ± 0.0063	**0.1369 ± 0.0163**	0.1458 ± 0.0102
Yeast	**0.2237 ± 0.0026**	0.2302 ± 0.0041	0.2278 ± 0.0029	0.2279 ± 0.0037	0.2332 ± 0.0044	0.2263 ± 0.0035	0.2319 ± 0.0042	0.2318 ± 0.003	0.2303 ± 0.0026
Enron	**0.0507 ± 0.0017**	0.0574 ± 0.0012	0.0516 ± 0.0013	0.0519 ± 0.0013	0.0532 ± 0.0012	0.055 ± 0.003	0.0508 ± 0.0014	0.0525 ± 0.0022	0.0512 ± 0.0019
Arts	**0.0609 ± 0.0009**	0.0614 ± 0.0007	0.0635 ± 0.0012	0.0644 ± 0.001	0.0634 ± 0.0007	0.0612 ± 0.0006	0.0622 ± 0.0009	0.0639 ± 0.0006	0.0612 ± 0.0008
Business	**0.0284 ± 0.0003**	0.0284 ± 0.0002	0.0293 ± 0.0005	0.0294 ± 0.0004	0.0292 ± 0.0004	0.0287 ± 0.0005	0.0291 ± 0.0005	0.0291 ± 0.0004	0.0288 ± 0.0003
Educations	0.0427 ± 0.0007	0.0436 ± 0.0007	0.0443 ± 0.0007	0.0445 ± 0.0008	0.0441 ± 0.001	0.0428 ± 0.0006	0.0431 ± 0.0007	0.044 ± 0.0006	**0.0423 ± 0.0007**
Entertain	**0.061 ± 0.0013**	0.0658 ± 0.0008	0.0657 ± 0.0013	0.0671 ± 0.0011	0.0659 ± 0.0014	0.0631 ± 0.0014	0.0654 ± 0.0011	0.0641 ± 0.001	0.0615 ± 0.0012
Recreation	**0.0605 ± 0.0008**	0.0619 ± 0.0012	0.0624 ± 0.0008	0.0648 ± 0.0007	0.0644 ± 0.0006	0.0613 ± 0.0011	0.0626 ± 0.0012	0.0651 ± 0.0007	0.061 ± 0.0011
Reference	**0.0311 ± 0.0005**	0.0313 ± 0.0012	0.0322 ± 0.0012	0.0336 ± 0.001	0.0329 ± 0.0002	0.0312 ± 0.0007	0.0321 ± 0.0009	0.0326 ± 0.0007	0.0315 ± 0.0006
Science	0.0349 ± 0.0004	0.0355 ± 0.0003	0.0358 ± 0.0004	0.0363 ± 0.0004	0.0358 ± 0.0004	0.0353 ± 0.0005	0.0355 ± 0.0005	0.0357 ± 0.0003	**0.0348 ± 0.0003**
Social	0.0269 ± 0.0007	0.0317 ± 0.0013	0.0303 ± 0.0005	0.0309 ± 0.0003	0.0287 ± 0.0007	0.0274 ± 0.0007	0.0282 ± 0.0006	0.0291 ± 0.0009	**0.0266 ± 0.0012**
Average	**0.0645**	0.0695	0.0677	0.0681	0.0706	0.0660	0.0687	0.0669	0.0663

器得出的结果，而 Macro-F1 指标则是基于 3NN 分类器的结果。所有算法均在选取数据集中 20%的特征子集上进行评估，并计算了平均分类性能及其标准偏差。表中以加粗字体表示的是在各个数据集上达到最优性能的特征选择方法。

Hamming Loss 的值越小，表明特征选择算法的分类性能越优良。如表 2 所示，所提出的算法在 8 个数据集上取得了最佳性能，在其他 4 个数据集上取得了次优性能。其中，在 Education、Science 和 Social 数据集上，AIII-FS 算法展现了最佳性能，这代表 AIII-FS 算法在这三个数据集上的表现结果要更好。相较于其他六种基于信息论的多标签特征选择算法(MIFS、D2F、PMU、SCLS、LRFS 和 FIMF)，所提出的算法在 12 个数据集上均展现出最佳性能。总体而言，本文提出的算法在 ML-KNN 分类器上实现了最佳的 Hamming Loss 分类性能。

Table 3. Comparison results of 9 multi-label feature selection algorithms on the Zero One Loss index
表 3. 9 个多标签特征选择算法在 Zero One Loss 指标上的比较结果

数据集	IPFS	MIFS	D2F	PMU	SCLS	LRFS	FIMF	FSSL	AIII-FS
Medical	**0.5008 ± 0.0395**	0.5468 ± 0.0826	0.6561 ± 0.0371	0.6626 ± 0.0407	0.8262 ± 0.0064	0.5774 ± 0.0383	0.5772 ± 0.0397	0.6205 ± 0.0983	0.8096 ± 0.0033
Scene	**0.5729 ± 0.0935**	0.8281 ± 0.1163	0.6087 ± 0.0642	0.611 ± 0.0666	0.7412 ± 0.0447	0.597 ± 0.0755	0.764 ± 0.0845	0.5807 ± 0.1115	0.634 ± 0.094
Yeast	0.8884 ± 0.0194	0.9266 ± 0.0401	**0.886 ± 0.0279**	0.8917 ± 0.0288	0.9167 ± 0.0118	0.8866 ± 0.0224	0.9045 ± 0.0263	0.9233 ± 0.0322	0.9158 ± 0.0292
Enron	0.8905 ± 0.0213	0.9817 ± 0.0064	0.8985 ± 0.0186	0.9036 ± 0.0286	0.9415 ± 0.0258	0.9264 ± 0.0368	0.8916 ± 0.0242	0.9055 ± 0.0266	**0.8897 ± 0.0319**
Arts	0.9149 ± 0.0238	0.9179 ± 0.029	0.9548 ± 0.0111	0.9706 ± 0.0168	0.9529 ± 0.0109	0.9198 ± 0.0272	0.9392 ± 0.0247	0.9636 ± 0.0219	**0.9093 ± 0.0244**
Business	0.47 ± 0.0076	**0.4651 ± 0.0054**	0.4809 ± 0.0116	0.4829 ± 0.0115	0.4763 ± 0.0107	0.4724 ± 0.0102	0.4788 ± 0.0126	0.4743 ± 0.012	0.473 ± 0.0086
Educations	**0.8885 ± 0.0311**	0.9486 ± 0.0299	0.9483 ± 0.0094	0.9549 ± 0.0143	0.9339 ± 0.0139	0.9017 ± 0.0196	0.9111 ± 0.0206	0.9418 ± 0.0268	0.8937 ± 0.0254
Entertain	**0.8202 ± 0.0416**	0.9303 ± 0.0279	0.9056 ± 0.0101	0.9414 ± 0.0087	0.9034 ± 0.0131	0.8574 ± 0.028	0.8922 ± 0.0309	0.8831 ± 0.0342	0.8344 ± 0.0363
Recreation	**0.8613 ± 0.0254**	0.8816 ± 0.0321	0.9207 ± 0.009	0.9712 ± 0.0061	0.9533 ± 0.0054	0.8782 ± 0.0204	0.8999 ± 0.0171	0.9568 ± 0.0219	0.8633 ± 0.0232
Reference	**0.6833 ± 0.0967**	0.7805 ± 0.0685	0.8031 ± 0.0381	0.8107 ± 0.0522	0.8284 ± 0.0373	0.7603 ± 0.0639	0.7559 ± 0.0619	0.8083 ± 0.0477	0.7224 ± 0.084
Science	0.9293 ± 0.0175	0.931 ± 0.0255	0.9725 ± 0.0058	0.9848 ± 0.0082	0.9549 ± 0.012	0.9403 ± 0.0118	0.9497 ± 0.0102	0.9514 ± 0.02	**0.9225 ± 0.0199**
Social	**0.6763 ± 0.0875**	0.8768 ± 0.0883	0.7324 ± 0.0875	0.775 ± 0.0686	0.7446 ± 0.0426	0.7216 ± 0.0516	0.7343 ± 0.0505	0.7514 ± 0.0808	0.7038 ± 0.0788
Average	**0.7580**	0.8345	0.8139	0.8300	0.8478	0.7865	0.8082	0.8134	0.7976

Zero One Loss 的值越小，代表分类性能越优。如表 3 所示，所提出的算法在 Medical、Scene、Education、Entertain、Recreation、Reference 和 Social 这七个数据集上展现了最优的分类性能，在 Enron、Arts、Business 和 Science 数据集上则次之，表现仅次于 AIII-FS 和 MIFS 算法。在 Yeast 数据集上，D2F 和 LRFS 算法的性能优于所提出的算法。从平均结果来看，所提出的算法在所有数据集上的表现均优于其他算法。

Macro-F1 值越高，表示分类性能越优良。根据表 4 所示的平均结果，在所有数据集上，提出算法 IPFS 平均值为 0.1691，而 LRFS 算法平均值为 0.1588；其他算法的平均值分别为 0.1194、0.1346、0.1244、

0.1372、0.1309、0.1377。显然，IPFS 在性能上显著优于这些算法。SCLS 算法取得了最低的平均值 0.1105。因此，IPFS 算法在 Macro-F1 指标上表现出明显的优越性。

Table 4. Comparison results of 9 multi-label feature selection algorithms on the Macro-F1 index
表 4. 9 个多标签特征选择算法在 Macro-F1 指标上的比较结果

数据集	IPFS	MIFS	D2F	PMU	SCLS	LRFS	FIMF	FSSL	AIII-FS
Medical	**0.2069 ± 0.0287**	0.161 ± 0.0209	0.1207 ± 0.0187	0.1138 ± 0.0178	0.0626 ± 0.006	0.1872 ± 0.0265	0.1856 ± 0.0275	0.1486 ± 0.0399	0.0402 ± 0.0018
Scene	0.5234 ± 0.0769	0.2882 ± 0.1416	0.4869 ± 0.0548	0.4934 ± 0.0709	0.3705 ± 0.0338	**0.5288 ± 0.0677**	0.3895 ± 0.0892	0.5247 ± 0.1187	0.4888 ± 0.0853
Yeast	0.3424 ± 0.045	0.282 ± 0.0589	**0.3476 ± 0.0388**	0.3402 ± 0.0307	0.3007 ± 0.0263	0.3426 ± 0.0413	0.3243 ± 0.0253	0.3091 ± 0.0419	0.3202 ± 0.0324
Enron	**0.133 ± 0.0183**	0.0873 ± 0.014	0.1243 ± 0.0144	0.1241 ± 0.0175	0.1106 ± 0.0127	0.1093 ± 0.022	0.1322 ± 0.0185	0.1182 ± 0.0214	0.1233 ± 0.0135
Arts	**0.1075 ± 0.0263**	0.0841 ± 0.0252	0.0638 ± 0.0103	0.0577 ± 0.0143	0.0721 ± 0.0158	0.0972 ± 0.0227	0.0845 ± 0.0251	0.0608 ± 0.0259	0.1004 ± 0.0242
Business	**0.0954 ± 0.0144**	0.0894 ± 0.0168	0.0687 ± 0.0058	0.0545 ± 0.0072	0.073 ± 0.0106	0.0839 ± 0.0081	0.0638 ± 0.009	0.0499 ± 0.0069	0.0881 ± 0.0101
Educations	**0.0825 ± 0.0145**	0.0446 ± 0.0205	0.0646 ± 0.009	0.0571 ± 0.009	0.0587 ± 0.0098	0.0742 ± 0.0128	0.0768 ± 0.0125	0.0628 ± 0.0148	0.0806 ± 0.0144
Entertain	**0.1411 ± 0.0245**	0.0814 ± 0.0244	0.1081 ± 0.0108	0.0833 ± 0.0138	0.0949 ± 0.0128	0.1359 ± 0.0218	0.1037 ± 0.0226	0.0938 ± 0.0284	0.14 ± 0.0211
Recreation	0.1239 ± 0.0233	0.1177 ± 0.0287	0.0824 ± 0.011	0.0526 ± 0.0103	0.0661 ± 0.0109	0.1199 ± 0.0211	0.0977 ± 0.0148	0.0571 ± 0.0266	**0.1287 ± 0.0223**
Reference	**0.0819 ± 0.0151**	0.072 ± 0.0148	0.0444 ± 0.0043	0.0341 ± 0.0071	0.0363 ± 0.0054	0.0709 ± 0.0115	0.0614 ± 0.01	0.0411 ± 0.0105	0.0752 ± 0.0115
Science	0.078 ± 0.0241	0.0601 ± 0.0197	0.0402 ± 0.0055	0.0284 ± 0.0078	0.0304 ± 0.0036	0.0632 ± 0.0143	0.0521 ± 0.0085	0.0391 ± 0.019	**0.0784 ± 0.0148**
Social	0.1139 ± 0.0191	0.0657 ± 0.0281	0.0642 ± 0.0063	0.0542 ± 0.006	0.0506 ± 0.0043	0.093 ± 0.0136	0.0757 ± 0.0117	0.0639 ± 0.0242	**0.1148 ± 0.0174**
Average	**0.1691**	0.1194	0.1346	0.1244	0.1105	0.1588	0.1372	0.1309	0.1377

为了更清晰地展示本文提出算法的优越性，图 2、图 3 分别给出了 9 个算法在 Arts 和 Education 数据集上的 20 组特征子集的平均分类性能，特征子集中的特征个数从总特征数的 1%到 20%。横坐标表示特征个数百分比，纵坐标表示不同评价标准的分类性能。不同颜色的线和符号代表不同的算法，其中 IPFS 算法用黑色菱形线表示。

从图 2 和图 3 中的结果可以观察到，在 Arts 数据集上，IPFS 的四个分类评价指标均展现出较其他多标签特征选择算法更优的分类性能。在 Education 数据集上，本研究提出的算法与 AIII-FS 算法在分类性能上表现相似，且均优于其他算法。综合上述分析，本研究提出的算法展现出较优的分类性能，这表明通过该算法选出的特征子集具有更高的质量。

5. 结语

本文提出了一种基于三元互信息的多标签特征选择算法 IPFS，该算法通过评估候选特征对标签集中成对标签的互信息贡献，重新定义了特征重要性的度量，并构建了相应的评估函数，解决了高维多标签

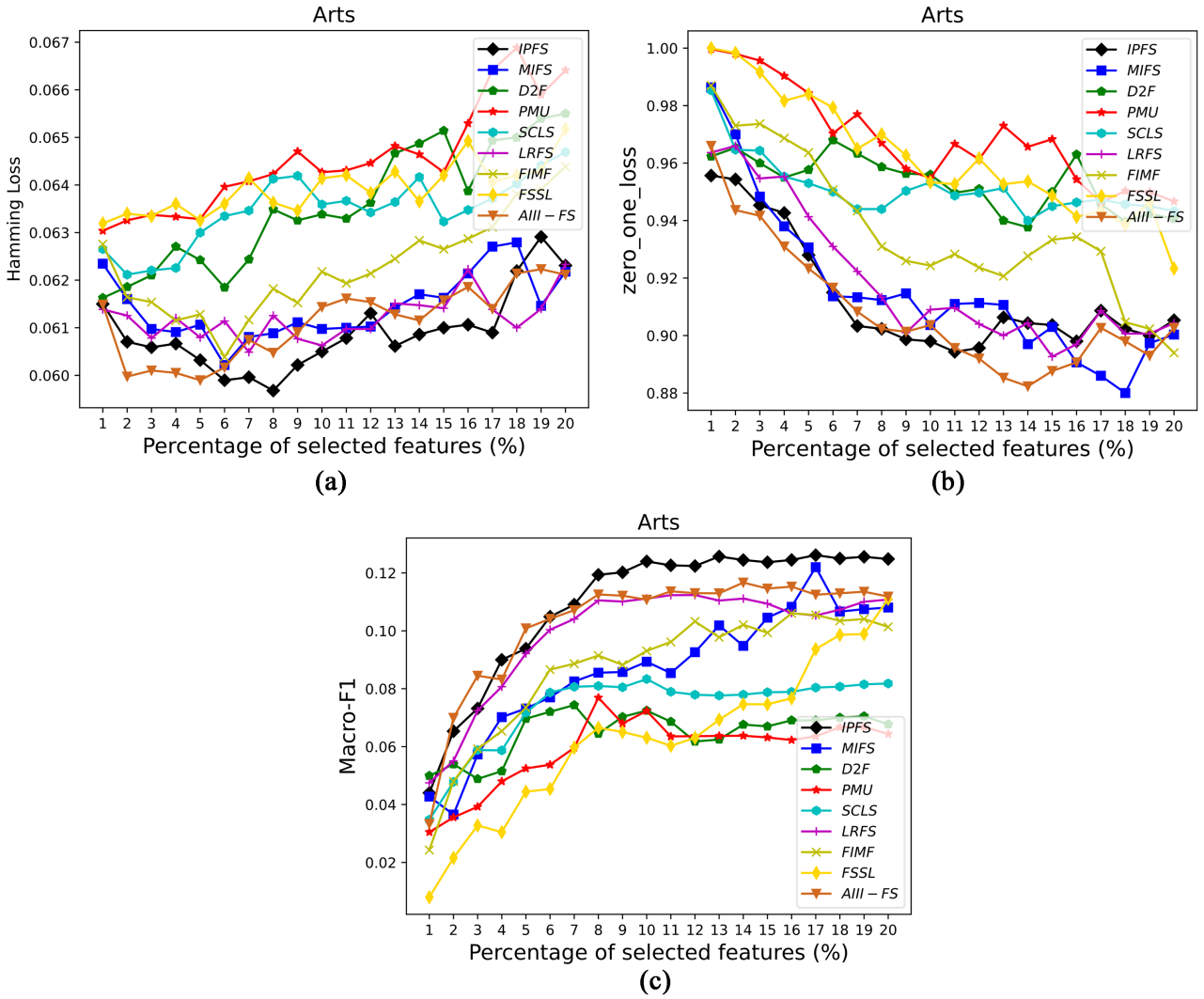
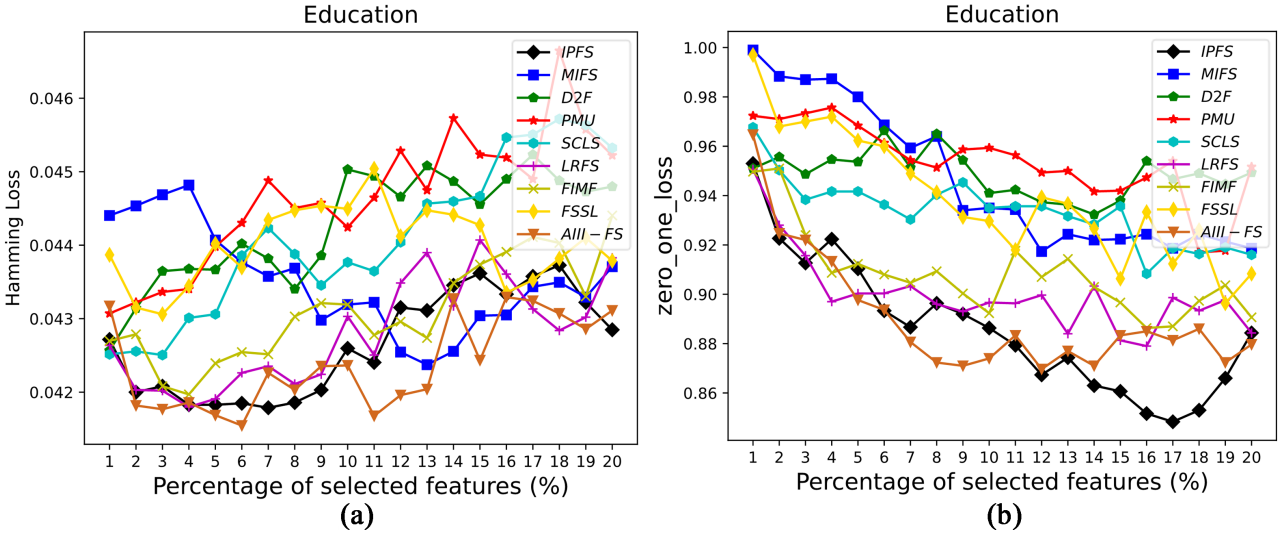


Figure 2. Experimental results of Arts dataset: (a) Hamming Loss, (b) Zero One Loss, (c) Macro-F1
图 2. Arts 数据集实验结果：(a) Hamming Loss, (b) Zero One Loss, (c) Macro-F1



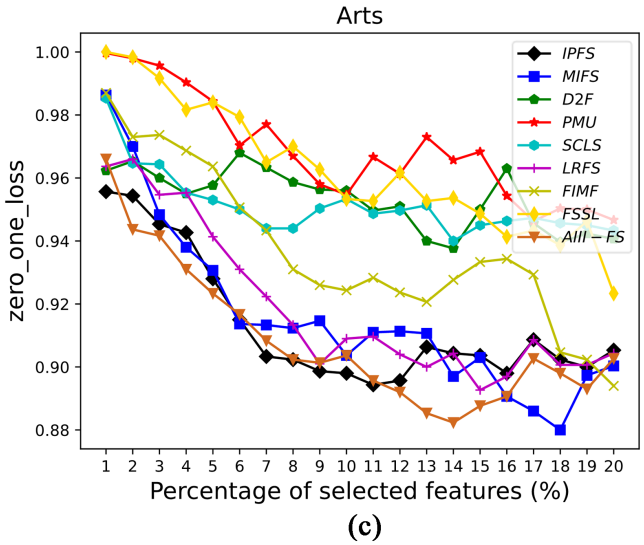


Figure 3. Experimental results of Education dataset: (a) Hamming Loss, (b) Zero One Loss, (c) Macro-F1
图 3. Education 数据集实验结果：(a) Hamming Loss, (b) Zero One Loss, (c) Macro-F1

数据存储与计算的挑战。利用前向搜索策略，该算法能够精确地评估特征与标签间的相关性，并筛选出与标签集合相关性最高且冗余性最低的特征子集。在 12 个公开数据集上的实验结果表明，该算法在 Hamming Loss 和 Macro-F1 等评价指标上的性能优于现有算法。该算法不仅减少了数据存储成本，增强了计算性能，还提高了多标签学习的分类性能，证明了其在多标签学习任务中的有效性和优越性。未来的研究可以进一步探索该算法在更广泛应用场景中的性能表现，并优化其计算效率与可扩展性。

基金项目

国家自然科学基金项目(62206085)；省部共建电工装备可靠性与智能化国家重点实验室(河北工业大学)优秀青年创新基金项目(EERI_OY2022005)；河北省省级科技计划项目(225676163GH)。

参考文献

[1] Papaspiliopoulos, O. (2020) High-Dimensional Probability: An Introduction with Applications in Data Science. *Quantitative Finance*, **20**, 1591-1594. https://doi.org/10.1080/14697688.2020.1813475

[2] 姜建武, 王博. 高维数据组合关联关系挖掘方法[J]. 科学技术与工程, 2023, 23(4): 1615-1624.

[3] Kundu, R. and Chattopadhyay, S. (2022) Deep Features Selection through Genetic Algorithm for Cervical Pre-Cancerous Cell Classification. *Multimedia Tools and Applications*, **82**, 13431-13452. https://doi.org/10.1007/s11042-022-13736-9

[4] Dutta, S. and Das, M. (2023) Remote Sensing Scene Classification under Scarcity of Labelled Samples—A Survey of the State-of-the-Arts. *Computers & Geosciences*, **171**, Article 105295. https://doi.org/10.1016/j.cageo.2022.105295

[5] Lee, J. and Kim, D. (2013) Feature Selection for Multi-Label Classification Using Multivariate Mutual Information. *Pattern Recognition Letters*, **34**, 349-357. https://doi.org/10.1016/j.patrec.2012.10.005

[6] Lee, J. and Kim, D. (2015) Mutual Information-Based Multi-Label Feature Selection Using Interaction Information. *Expert Systems with Applications*, **42**, 2013-2025. https://doi.org/10.1016/j.eswa.2014.09.063

[7] Lee, J. and Kim, D. (2017) SCLS: Multi-Label Feature Selection Based on Scalable Criterion for Large Label Set. *Pattern Recognition*, **66**, 342-352. https://doi.org/10.1016/j.patcog.2017.01.014

[8] Jian, L., Li, J., Shu, K., *et al*. (2016) Multi-Label Informed Feature Selection. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, New York, 9-15 July 2016, 1627-1633.

[9] Lee, J. and Kim, D. (2015) Fast Multi-Label Feature Selection Based on Information-Theoretic Feature Ranking. *Pattern Recognition*, **48**, 2761-2771. https://doi.org/10.1016/j.patcog.2015.04.009

[10] Zhang, P., Liu, G. and Gao, W. (2019) Distinguishing Two Types of Labels for Multi-Label Feature Selection. *Pattern*

Recognition, **95**, 72-82. https://doi.org/10.1016/j.patcog.2019.06.004

[11] Liu, J., Li, Y., Weng, W., *et al*. (2020) Feature Selection for Multi-Label Learning with Streaming Label. *Neurocomputing*, **387**, 268-278.

[12] Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[13] Pan, M., Sun, Z., Wang, C. and Cao, G. (2022) A Multi-Label Feature Selection Method Based on an Approximation of Interaction Information. *Intelligent Data Analysis*, **26**, 823-840. https://doi.org/10.3233/ida-215985

[14] Grigorios, T., Eleftherios, S.-X. and Jozef, V. (2011) Mulan: A Java Library for Multi-Label Learning. *Journal of Machine Learning Research*, **12**, 2411-2414.

[15] Zhang, M. and Zhou, Z. (2007) ML-KNN: A Lazy Learning Approach to Multi-Label Learning. *Pattern Recognition*, **40**, 2038-2048. https://doi.org/10.1016/j.patcog.2006.12.019