

# 基于最大化联合互信息和最小化联合熵的特征选择

刘亚文, 温 勇

南京邮电大学理学院, 江苏 南京

收稿日期: 2023年3月11日; 录用日期: 2023年4月7日; 发布日期: 2023年4月13日

## 摘 要

随着大数据时代的到来, 数据越来越容易获得, 同时获得的数据的维度也越来越高。高维的数据能够详细记录事物的属性, 但维度越高, 冗余数据也就越多。从数据中剔除冗余特征就显得很重要。基于互信息(MI)的特征选择方法能够有效地降低数据维数和提高分类精度。但现有的方法在特征选择的过程中评判特征的标准单一, 无法有效排除冗余特征。本文为此提出了一种基于最大联合互信息和最小联合熵的特征选择方法(JMIMJE)。JMIMJE在特征选择时考虑了整体联合互信息和联合熵两种因素, 联合互信息衡量特征子集整体的与分类的相关性, 联合熵衡量特征子集的稳定性和预测精度。JMIMJE在筛选特征时对特征子集的相关性和稳定性进行了平衡。在预测精度方面, JMIMJE比mRMR (最小冗余度最大相关性)提高了2个百分点; 与联合互信息(JMI)相比提高了1个百分点。

## 关键词

信息熵, 互信息, 联合互信息, 联合熵, 特征选择, 降维

# Feature Selection Based on Maximizing Joint Mutual Information and Minimizing Joint Entropy

Yawen Liu, Yong Wen

College of Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Mar. 11<sup>th</sup>, 2023; accepted: Apr. 7<sup>th</sup>, 2023; published: Apr. 13<sup>th</sup>, 2023

## Abstract

With the advent of the data era, data is getting easier and easier to obtain, and the dimensions of

the data are getting higher and higher. Higher-dimensional data can record the attributes of things in detail, but the higher the dimension, the more redundant data. It is important to remove redundant features from the data. The feature selection method based on mutual information (MI) is not good at reducing the data dimension and improving the classification accuracy. In the process of feature selection, the existing methods have a single feature evaluation criterion and can not effectively eliminate redundant features. A feature selection method (JMIMJE) based on maximum joint mutual information and minimum joint entropy is proposed. JMIMJE considers two factors, global joint mutual information and joint entropy, in feature selection. Combined mutual information to measure the correlation between the whole feature subset and the classification, combined entropy to measure the stability of the feature subset. JMIMJE balances the correlation and stability of feature subsets during feature screening. In terms of prediction accuracy, JMIMJE is 2 percentage points higher than mRMR (minimum redundancy maximum correlation). Compared with Joint Mutual Information (JMI), an increase of 1 percentage point.

## Keywords

Information Entropy, Mutual Information, Joint Mutual Information, Joint Entropy, Feature Selection, Dimension Reduction

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着大数据时代的到来,我们能够得到目标的属性信息越来越多,同时也会得到大量的冗余信息。在对数据进行分析之前需要对数据进行筛选。特征选择方法是从高维数据中筛选出相关特征,减少特征个数,从而降低数据分析的复杂度和提高数据的解释度。互信息的优点是计算简单、可解释性强。互信息在深度学习、特征选择中有着广泛的应用。互信息特征选择的一个方向是 Filter [1]类型特征选择的一个方向。目前互信息特征选择方法主要从互信息、条件互信息、联合互信息、联合条件互信息方面研究的,其中条件互信息、联合互信息和联合条件互信息的计算都比较繁琐。为了不再降低特征选择的效率,一般使用选择路线复杂度低、效率高的贪婪向前搜索法。

为了解决互信息无法对相似特征评价的问题,董泽民等[2]提出了一种基于联合互信息(joint mutual information, JMI)的特征选择方法。JMI 特征选择方法在筛选时不仅考虑特征子集与分类之间的关系,还考虑特征子集与未选特征子集之间的相关性。联合互信息特征选择方法在筛选特征时,随着筛选特征个数的增加,不同未选特征和已选特征的联合互信息趋近。当筛选特征变多时, JMI 筛选效果就会变得不佳。本文提出基于最大化联合互信息和最小化联合熵的特征选择方法(feature selection based on maximizing joint mutual information and minimizing joint entropy, JMIMJE)。JMIMJE 方法利用联合互信息筛选出效果明显的特征组合,再利用联合熵对联合互信息筛选出的特征组合进行二次筛选,这样筛选出的特征子集即可以保证特征子集与整体的相关性,又可以保证特征子集的稳定性的。

## 2. 相关工作

最早应用于特征选择的互信息是信息增益(Information Gain, IG) [3],信息增益只根据特征与分类之间的互信息大小对特征进行筛选,无法对特征是否冗余进行判断。Liu [4]等人提出了一种可以判断冗余特

征的信息增益文本方法。Battiti 等人[5]提出了不再假设特征之间相互独立的特征选择方法(Mutual Information Feature Selection, MIFS), 但 MIFS 仍有不足之处。Fleuret 等[6]提出了通过条件互信息降低数据维数的特征选择方法。

Hoque 等[7]和 Cho 等[8]提出了 MIFS 的改进方法, MIFS-ND 方法和归一化互信息特征选择(Normalized Mutual Information Feature Selection, NMIFS)方法。这两种方法的效果都比 MIFS 的效果表现的好, 但标准不一致的问题没有解决。Peng 等[9]提出了由 MIFS 方法改进而来的基于互信息的最小冗余度最大相关性(minimum Redundancy Maximum Relevance, mRMR)的特征选择方法, 但 mRMR 方法对相似特征判断能力较差。

为了解决对相似特征判断的问题, 董泽民等提出了基于联合互信息(Joint Mutual Information, JMI)的特征选择方法。Bennasar 等[10]在考虑联合互信息整体稳定性之后提出了基于最大化联合互信息(Joint Mutual Information Maximisation, JMIM)的特征选择方法。

综上在进行特征选择时, 判断特征是否冗余, 筛选出的特征子集会因特征评价函数不同而不同。本文采用联合互信息和联合熵结合的方式, 判断特征子集与分类相关性和特征子集的稳定性, 希望能提高筛选特征的效果。

### 3. 互信息理论

1) 信息熵: 1948 年, 英国数学家香农提出了“信息熵”概念。熵是随机变量的不确定度的度量, 不确定程度越大, 信息熵就越大, 信息熵越大概率就越小。

信息熵的公式为:

$$H(X) = - \sum_{x_i \in X} p(x_i) \lg(p(x_i)) \quad (1)$$

式中,  $P(x_i)$ 表示的是  $x_i$  在  $X$  中的概率密度。

2) 联合熵: 联合熵就是度量一个联合分布的随机系统的不确定度。两个变量的联合熵表示两个变量在一起的不确定度。

联合熵的公式为:

$$H(X, C) = - \sum_{x_i \in X} \sum_{c_j \in C} p(x_i, c_j) \lg(p(x_i, c_j)) \quad (2)$$

式中,  $H(X, C)$ 越大, 说明随机变量  $X$  和  $C$  的联合分布的随机系统的不确定度就越大,  $P(x_i, c_j)$ 表示两个变量的组合概率密度函数。

3) 互信息: 互信息是一个变量与另一个变量的共享信息的度量。

互信息公式为:

$$I(X; C) = \sum_{x_i \in X} \sum_{c_j \in C} p(x_i, c_j) \lg \left( \frac{p(x_i, c_j)}{p(x_i)p(c_j)} \right) \quad (3)$$

式中,  $I(X, C)$ 表示  $X$  与  $C$  之间的共享信息度量,  $I(X, C)$ 越大说明  $X$  与  $C$  之间的相关性越强。

4) 联合互信息: 联合互信息是一个联合分布的随机系统整体包含一个随机变量的信息量。

联合互信息公式为:

$$I(X, Y; C) = \sum_{x_i \in X} \sum_{y_j \in Y} \sum_{c_k \in C} p(x_i, y_j, c_k) \lg \left( \frac{p(x_i, y_j, c_k)}{p(x_i)p(y_j)p(c_k)} \right) \quad (4)$$

式中,  $I(X;C)$ 表示  $X, Y$  整体与  $C$  之间的共享信息度量,  $I(X,Y;C)$ 越大说明  $X, Y$  整体与  $C$  之间的相关性越强。

#### 4. 问题陈述与分析

对高维数据进行筛选时, 我们总是尽可能的剔除冗余的特征, 利用少量的数据进行分析, 对少量数据进行分析得到的结果具有吸引力和说服力。

设  $X = (x_1, x_2, \dots, x_n)$ ,  $X$  为全特征集合,  $x_i \in X$ ,  $1 \leq i \leq n$ ,  $n$  表示为特征总量;  $X = (x_1^i, x_2^i, \dots, x_m^i)$  表示为所选特征集合的一个样本的样本值,  $1 \leq m \leq n$ 。当  $n$  较大, 且  $m \ll n$  时, 不同的特征评价标准筛选出的特征子集不同, 这时用互信息和联合互信息来计算, 效果不佳[11]。

基于互信息的特征选择是从全部特征集合中挑选出  $w$  个特征构成的特征子集  $S \in (x_1, x_2, \dots, x_w)$ ,  $S \in X$ , 在高维多特征数据子集的特征筛选时, 特征子集的联合互信息是小于因变量的信息熵。特征子集的联合互信息是对因变量信息熵的解释, 若联合互信息 = 因变量的信息熵, 那么特征子集就能百分百解释因变量。特征子集的联合熵是特征子集的不确定度。在高维数据中, 联合互信息随着特征子集个数的增加而增加, 联合互信息的增加速度是递减的。随着特征子集中特征个数的增加, 待选特征子集的联合互信息比较接近, 这时用联合互信息筛选特征子集的效果不好; 在联合互信息相近时, 联合熵小的特征子集的效果好。

下面对不同特征选择方法的计算过程进行比较。

##### 1) IG 特征选择:

这种方法是将每个特征与分类的信息增益值从大到小排序, 选取前  $k(1 \leq k \leq n)$  特征即可。IG 只能辨别单个特征对分类的相关性, 无法衡量多个特征对分类的相关性。

##### 2) mRMR 特征选择:

$$\alpha(f_i) = I(f_i; C) - \frac{1}{S} \sum_{f_s \in S} I(f_i; f_s) \quad (5)$$

其中计算以  $I(f_i; C)$  为基础值, 考虑未加入的特征与每个子特征的互信息。mRMR 特征选择方法能够排除与已选特征具有直接相关的特征, 但无法筛选具有间接相关的特征, 当  $f_i \in F - S$  时,  $f_{s1}, f_{s2} \in S$ ,  $f_i$  分别与  $f_{s1}, f_{s2}$  独立, 当  $I(f_i; f_{s1})$  与  $I(f_i; f_{s2})$  接近时, mRMR 无法判断新加入的特征  $f_i$  是否合适。mRMR 特征选择能够排除与已选特征具有直接相关的特征, 但无法筛选具有间接相关的特征。当用 JMI 筛选特征时就不会出现这样的问题。

##### 3) JMI 特征选择:

$$\alpha(f_i) = \sum_{f_s \in S} I(f_i, f_s; C) \quad (6)$$

根据 JMI 的计算公式可以得出, JMI 考虑每一个  $f_i$  加入之后的特征子集整体与分类的联合互信息。JMI 在筛选特征时考虑的是特征子集与分类的相关性, 只要  $\sum_{f_s \in S} I(f_i, f_s; C)$  计算值较大即可。

可以通过表 1 的 ionosphere 的数据集说明 JMI 存在的问题。ionosphere 数据集是离散 + 连续型数据集, 该数据集共有 34 个特征和一个分类。JMI 在筛选特征时, 随着筛选特征个数的增加, 筛选特征子集的稳定性的就更加重要。表 2 为 JMI 筛选特征的互信息和联合互信息。在筛选特征子集的联合互信息接近时, 特征子集的稳定性与特征子集分类的准确性成正比。

该数据集共有 34 个特征  $(f_1, f_2, \dots, f_{34})$  和分类  $f_{35}$ , 样本数量为 351 个。根据 JMI 计算方式筛选的前三个特征为  $f_{14}, f_{19}, f_{15}$ , 其联合互信息和联合熵为 0.9273517、5.636871;  $f_{14}, f_{19}, f_3$  的联合互信息和联合熵为 0.9140590、4.376591。在分类的准确率上  $f_{14}, f_{19}, f_{15}$  的准确率小于  $f_{14}, f_{19}, f_3$  的准确率。在联合互信息接近时, 联合熵小的特征子集的分类准确率高于联合熵大的特征子集。

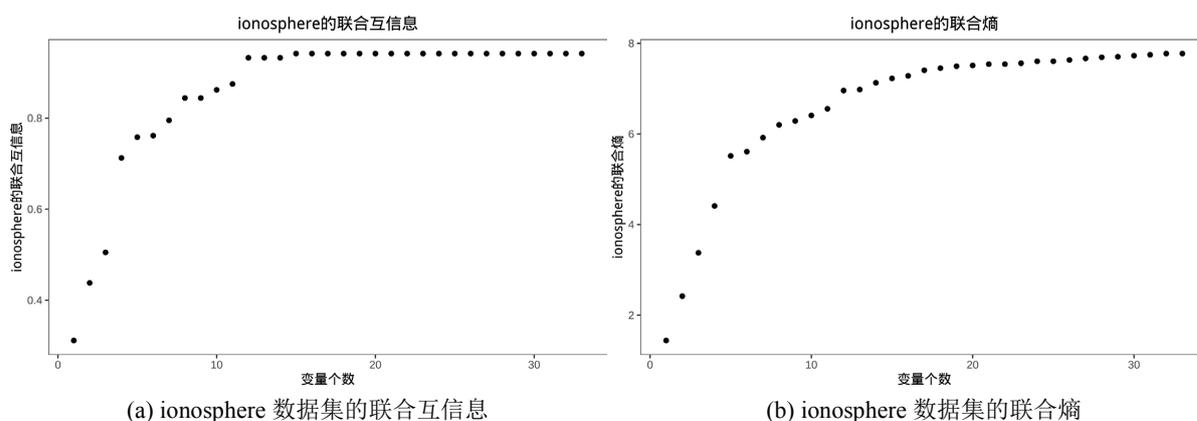
**Table 1.** Characteristics and classification of ionosphere data  
**表 1.** Ionosphere 数据的特征及分类

样本	$f_1$	$f_2$	$f_3$	...	$f_{34}$	$f_{35}$
1	1	0	0.99539	...	-0.45300	g
2	1	0	1	...	-0.02447	b
3	1	0	1	...	-0.38238	g
...	...	...	...	...	...	...
350	1	0	0.90608	...	-0.16243	g
351	1	0	0.84710	...	-0.06151	g

**Table 2.** Ionosphere mutual information and joint mutual information  
**表 2.** Ionosphere 的互信息和联合互信息

互信息	联合互信息
$I(f_1; C) = 0.17759734$	$I(f_{14}, f_{15}; C) = 0.4554273$
$I(f_2; C) = 0$	$I(f_{14}, f_2; C) = 0.3115936$
...	...
$I(f_{33}; C) = 0.13556091$	$I(f_{14}, f_{34}; C) = 0.4027734$
$I(f_{34}; C) = 0.04688331$	

通过对 IG、mRMR、JMI 方法分析可以得出, mRMR 在进行特征选择时考虑单个特征加入后的互信息的大小, 无法排除两个及两个以上特征相关情况。JMI 通过联合互信息的方式降低了两个及以上特征相关情况, 无法排除稳定性较差的特征子集。随着已选特征子集  $S$  中特征个数的增加, 特征子集的稳定性也就显得重要起来。从下图 1 可以得出随着特征子集  $S$  个数的增加, ionosphere 的联合互信息增加速度呈指数递减, 当选择的特征子集个数达到  $g$  ( $0 < g \leq n$ ) 时, 该数据集的联合互信息稳定下来。ionosphere 的联合熵增加速度也呈指数递减, 但减小幅度小于联合互信息的减小幅度。



**Figure 1.** Joint mutual information and joint entropy of the ionosphere dataset  
**图 1.** Ionosphere 数据集的联合互信息和联合熵

对于以上方法存在的问题, 本文提出的方法兼顾特征子集的相关性和稳定性。当有多个相似值存在时, 通过联合熵的选择在筛选特征子集与分类的相关性和筛选特征子集的稳定性之间找到平衡。

## 5. JMIMJE 总体思路

JMIMJE 特征选择方法从联合互信息和联合熵两方面筛选特征, 通过对已选特征与未选特征的联合互信息筛选出联合互信息较大值的特征集合  $C$ , 再通过联合熵对特征集合  $C$  进行二次筛选。这种方法可以解决现有联合互信息在特征选择过程中出现的无法排除冗余及不相关特征的问题。在指定子集大小的情况下, 挑选出的特征子集  $S$  与分类的相关性最大, 稳定性较高。

### 5.1. 最大联合互信息

1) 特征相关性: 当已选特征子集  $S = \{f_1, f_2, \dots, f_{s-1}, f_s\}$ ,  $f_i, f_j \in F - S$ , 如果  $I(f_i, S; C) > I(f_j, S; C)$ , 那么在已选特征子集  $S$  的前提下, 特征  $f_i$  和已选特征子集  $S$  整体与分类  $C$  的相关性强于特征  $f_j$  和已选特征子集  $S$  整体与分类  $C$  的相关性。

2) 最大联合互信息的计算:

**Table 3.** Maximum joint mutual information schematic table

**表 3.** 最大联合互信息示意图

$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
null	$I(f_1, f_2; C)$	$I(f_1, f_3; C)$	$I(f_1, f_4; C)$	$I(f_1, f_5; C)$
null	null	$I(f_1, f_2, f_3; C)$	$I(f_1, f_2, f_4; C)$	$I(f_1, f_2, f_5; C)$

表 3 中, 当子集  $S = \{f_1\}$  时, 从未选特征集合  $\{f_2, f_3, f_4, f_5\} \in F - S$  中筛选特征。表中的 null 表示已加入集合  $S$  的特征, 将  $\{f_2, f_3, f_4, f_5\}$  分别与子集  $\{f_1\}$  计算得到  $I(f_1, f_2; C)$ 、 $I(f_1, f_3; C)$ 、 $I(f_1, f_4; C)$  和  $I(f_1, f_5; C)$ 。当  $I(f_1, f_2; C)$  为最大值时, 从  $F - S$  中筛选出最佳特征  $f_2$ , 此时筛选出的特征集合  $S = \{f_1, f_2\}$  为最佳特征子集。同理, 从  $\{f_3, f_4, f_5\}$  挑选出联合互信息最大值加入集合  $S$  中, 这样筛选出的特征子集  $S$  与  $C$  的相关性较强。

### 5.2. 最小联合熵

**Table 4.** Minimum joint entropy schematic table

**表 4.** 最小联合熵示意图

$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
null	$H(f_1, f_2)$	$H(f_1, f_3)$	$H(f_1, f_4)$	$H(f_1, f_5)$
null	null	$H(f_1, f_2, f_3)$	$H(f_1, f_2, f_4)$	$H(f_1, f_2, f_5)$

表 4 中, 当子集  $S = \{f_1\}$ , 将  $\{f_2, f_3, f_4, f_5\}$  分别与子集  $\{f_1\}$  计算得到  $H(f_1, f_2)$ 、 $H(f_1, f_3)$ 、 $H(f_1, f_4)$  和  $H(f_1, f_5)$ , 当  $H(f_1, f_2)$  为最小值时, 从  $F - S$  中筛选出特征  $f_2$ , 说明特征子集  $\{f_1, f_2\}$  的稳定性最好。同理, 从  $\{f_3, f_4, f_5\}$  筛选出联合熵最小值加入集合  $S$  中, 这样筛选出的集合  $S$  的稳定性较强。

### 5.3. 最大联合互信息和最小联合熵

在特征选择的过程中, 我们希望每次选择的特征集合  $S$  与分类  $C$  之间的相关性都能得到最大提升, 联合互信息值体现了特征子集整体与分类的相关性, 联合熵体现了特征子集的不确定度。本文希望得到的是与分类相关性强和稳定性强的特征子集, 对这样的子集进行回归、分类和分析时得到的结果就更准确。

本文先用联合互信息的值筛选出未选特征集合  $\{F-S\}$  中联合互信息最大值的特征  $i$ , 或者与联合互信息最大值接近的特征  $j, k$ ; 然后再计算出未选特征  $\{i, j, k\}$  与已选特征集合  $S$  的联合熵, 未选特征  $\{i, j, k\}$  与已选特征集合  $S$  的联合互信息为  $M$ , 与已选特征集合  $S$  的联合熵为  $H$ ; 选出特征  $\{i, j, k\}$  中  $M/H$  值最大的特征加入特征集合  $S$  中。这样选出的特征集合与分类  $C$  之间的相关性能得到保证, 其稳定性也强于其他方法筛选的特征子集。

#### 5.4. 方法流程

输入: 数据集  $F$ , 分类  $C$ , 特征数  $N$ , 要选择的特征个数  $num$ ;

输出: 要选择的特征集合  $S$

- 1)  $S \leftarrow \emptyset$
- 2)  $list \leftarrow 0$ /计算特征与分类的互信息
- 3)  $S = \max(list)$ /选取最大的互信息特征加入  $S$
- 4) For  $i = 1$  to  $(N-S)$  do
- 5) 计算  $F-S$  中每个特征和已选特征子集  $S$  与分类  $C$  的联合互信息并储存  $M$
- 6) End for
- 7) 选取  $M$  中联合互信息最大值的特征, 加入到特征子集  $S$  中
- 8) 重复步骤 4~7, 直到选取  $num-2$  个特征
- 9) 从  $F-S$  中选取 2 个特征计算联合互信息并储存  $M$
- 10) 选取  $M$  中联合互信息最大值和联合互信息最大值近似的特征
- 11) 计算选取的特征的联合熵  $H$
- 12) 选取  $M/H$  值最大的特征加入特征子集  $S$
- 13) 输出特征子集  $S$

### 6. 实验验证

#### 6.1. 实验方案

表 5 的实验数据来自 UCI 公开的数据集, 实验数据来自 UCI 公开的数据集, 实验数据包括离散数据、连续数据、离散 + 连续数据。实验中还将 IG、mRMR、JMI 与 JMIMJE 进行比较, 用判别分析对选择的特征进行验证。通过分类的正确率判断选择特征的合理性。

连续变量的联合互信息的计算是比较繁琐的。目前, 连续变量联合互信息的计算方法是将连续变量离散化, 然后计算离散变量的联合互信息。连续变量的联合互信息与连续变量离散化的联合互信息还是有一些区别的。将每个数据集随机划分出 70% 的数据用作训练集, 整个数据集用作测试集; 将实验数据中的连续型数据转化成离散型数据; 再将不同方法已选的特征子集  $S$  输入到贝叶斯分类模型中, 用分类的准确率来判断特征子集的合理性。图 2 为实验流程图。

Table 5. Experimental data set

表 5. 实验数据集

编号	数据集	样本数	特征	数据类型
1	wine	178	14	连续
2	ionosphere	351	35	离散 + 连续
3	wdbc	569	32	离散 + 连续
4	kr-vs-kp	3196	37	离散

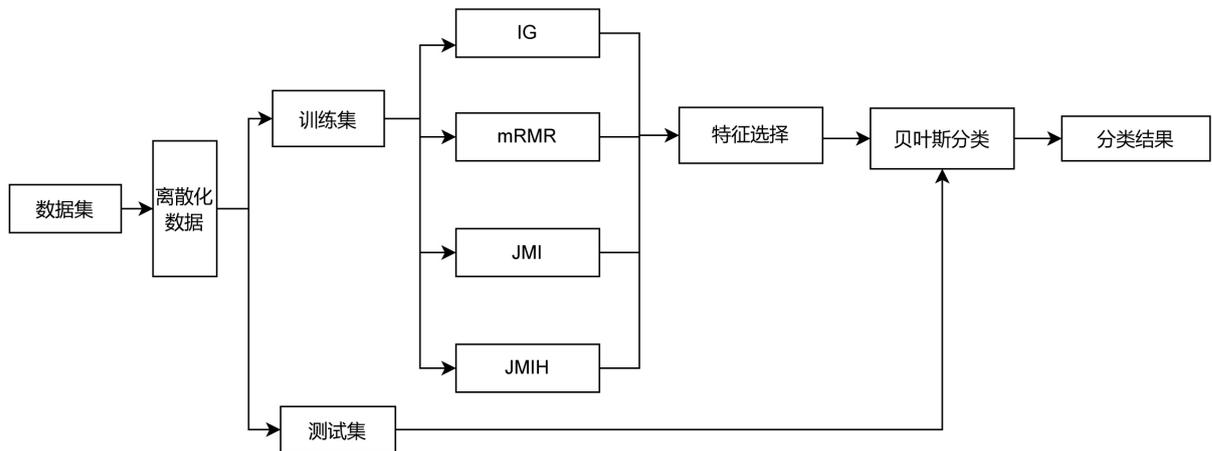


Figure 2. Experimental flow  
图 2. 实验流程

### 6.2. 实验预测精度分析

图 3 中, 横坐标表示数据集上选择的特征个数, 纵坐标表示对应特征个数下贝叶斯分类预测精度。通过不同的数据集上不同特征选择方法的分类预测精度变化图可以得出下列结论。

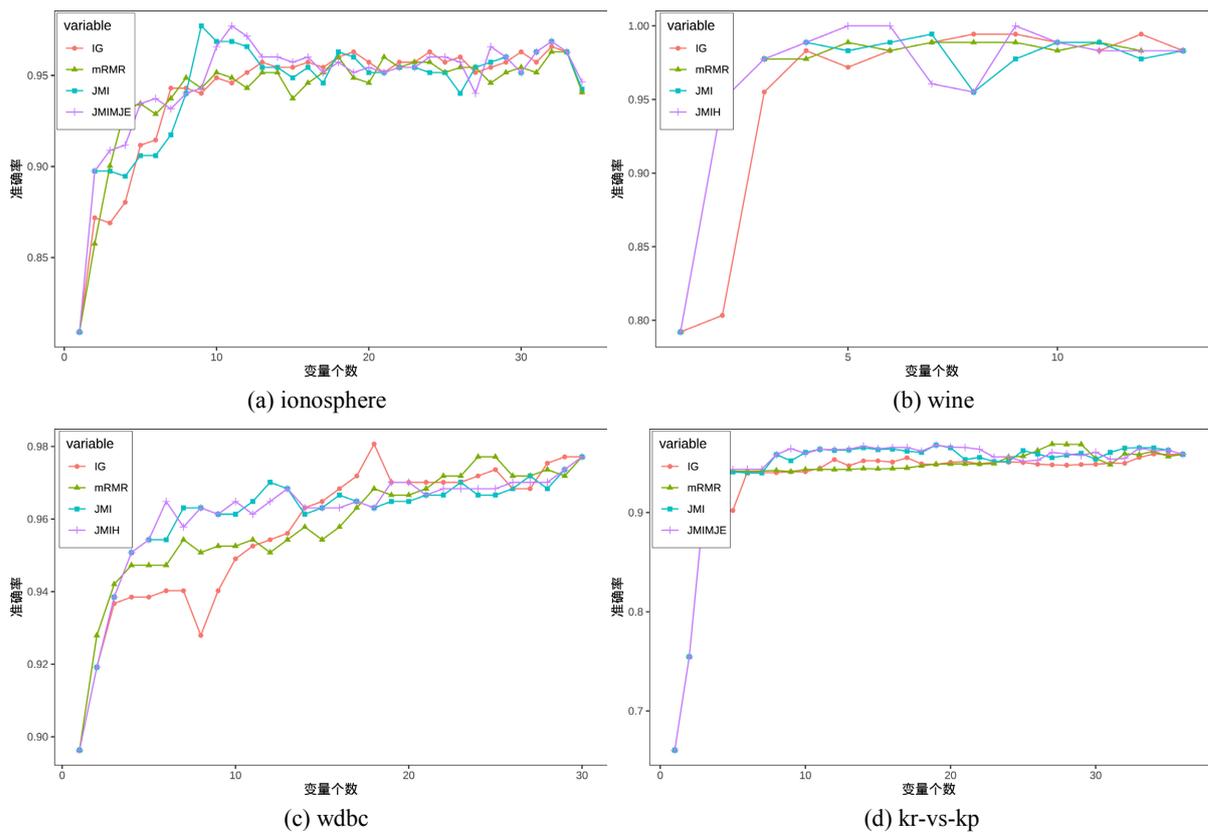


Figure 3. Different methods are accurate in selecting the number of features  
图 3. 不同方法在选择特征个数是精确度

1) 从图 3 中可以看到 JMIMJE 在样本较少的数据集上的预测结果: 在 ionosphere 数据集中, JMIMJE

在筛选第 9、10 个特征时的分类准确率小于 JMI 筛选特征的分类准确率; wine 数据集用 JMIMJE 筛选特征第 7、8、11、12 个特征时的分类准确率小于 JMI 的分类准确率。在预测精度上, JMIMJE 的分类准确率要高于 JMI 的分类准确率; kr-vs-kp 数据集用 JMIMJE 筛选少量特征(小于 50%特征总数)时的分类准确率优于其他三种特征选择方法的分类准确率。

2) 在特征个数逐渐增加时, 特征选择方法在预测精度上呈现先增加后减小的趋势: IG、mRMR 在特征选择时并没有明显表现出这一趋势。IG 和 mRMR 的预测精度有上下波动的情况, 这是 IG、mRMR 在特征选择时考虑的标准比较单一所造成的, 加入的特征为提供分类的信息量相对于其他变量较少。

3) 表 6 的 wine 数据集上, 在 wine 数据集上, JMIMJE 的平均预测精度比 IG、mRMR 的平均预测精度提升了 1 个百分点, 比 JMI 提升了 0.1 个百分点。在 ionosphere 数据集上, JMI 的平均预测精度最低, 原因是在筛选少量特征的预测精度时出现了较大的偏差。在 wdbc 数据集上, JMIMJE 的预测精度最高。在 kr-vs-kp 数据集上, JMIMJE 的平均预测精度最高, 其预测精度在少量(低于 50%)特征上的预测精度明显高于其他三种方法。

4) JMIMJE 特征选择方法在分类时考虑了特征子集与分类的相关性和特征子集的稳定性。JMI 和 JMIMJE 的预测精度波动总体上小于 IG、mRMR 的预测精度的波动, 这说明了 JMI 和 JMIMJE 的预测稳定性好于 IG 和 mRMR。在预测精度上, JMIMJE 的平均预测精度要高于 IG、mRMR 和 JMI。

**Table 6.** Comparison of prediction accuracy of different methods

**表 6.** 不同方法的预测精度对比

数据集	IG	mRMR	JMI	JMIMJE
wine	0.95271535	0.965823958	0.963483133	0.964887642
ionosphere	0.939280622	0.939280622	0.935897434	0.939725784
wdbc	0.95670767	0.958172233	0.96033978	0.960808433
kr-vs-kp	0.9450817	0.9468775	0.9493396	0.9495606

## 7. 结论

本文比较了 4 种特征选择方法筛选特征在不同数据集上的平均预测精度, JMIMJE 从联合互信息和联合熵两个方面对特征选择进行筛选, 在保证特征子集与分类相关性的同时增加了特征子集的稳定性。JMIMJE 平均预测精度高于 IG、mRMR 和 JMI 的平均预测精度, 其波动性小于 IG、mRMR 和 JMI 的预测波动性。同时, JMIMJE 也有其不足之处, 在数据不均衡的情况下筛选效果不佳。

## 参考文献

- [1] Gandhi, S.S. and Prabhune, S.S. (2017) Overview of Feature Subset Selection Algorithm for High Dimensional Data. *ICISC 2017: Proceedings of the 2017 IEEE International Conference on Inventive Systems and Control*, Coimbatore, 19-20 January 2017, 1-6. <https://doi.org/10.1109/ICISC.2017.8068599>
- [2] Fleuret, F. (2004) Fast Binary Feature Selection with Conditional Mutual Information. *Journal of Machine Learning Research*, **5**, 1531-1555.
- [3] 董泽民, 石强. 基于归一化模糊联合互信息最大的特征选择[J]. *计算机工程与应用*, 2017, 53(22): 105-110.
- [4] 黄志艳. 一种基于信息增益的特征选择方法[J]. *山东农业大学学报(自然科学版)*, 2013, 44(2): 252-256.
- [5] Liu, C., Wang, W., Zhao, Q., et al. (2017) A New Feature Selection Method Based on a Validity Index of Feature Subset. *Pattern Recognition Letters*, **92**, 1-8. <https://doi.org/10.1016/j.patrec.2017.03.018>
- [6] Battiti, R. (1994) Using Mutual Information for Selecting Features in Supervised Neural Net Learning. *IEEE Transactions on Neural Networks*, **5**, 537-550. <https://doi.org/10.1109/72.298224>

- [7] Hoque, N., Bhattacharyya, D.K. and Kalita, J.K. (2014) MIFS-ND: A Mutual Information-Based Feature Selection Method. *Expert Systems with Applications*, **41**, 6371-6385. <https://doi.org/10.1016/j.eswa.2014.04.019>
- [8] Cho, D. and Lee, B. (2017) Optimized Automatic Sleep Stage Classification Using the Normalized Mutual Information Feature Selection (NMIFS) Method. *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Jeju, 11-15 July 2017, 3094-3097. <https://doi.org/10.1109/EMBC.2017.8037511>
- [9] Peng, H., Long, F. and Ding, C. (2005) Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1226-1238. <https://doi.org/10.1109/TPAMI.2005.159>
- [10] Bennisarm, H.Y. and Setchi, R. (2015) Feature Selection Using Joint Mutual Information Maximisation. *Expert Systems with Applications*, **42**, 8520-8532. <https://doi.org/10.1016/j.eswa.2015.07.007>
- [11] Amaratunga, D. and Cabrera, J. (2016) High-Dimensional Data. *Journal of the National Science Foundation of Sri Lanka*, **44**, 3-9. <https://doi.org/10.4038/jnsfsr.v44i1.7976>