

# 基于信息熵的半监督特征选择算法

王 锋 刘吉超 魏 巍

(山西大学计算机与信息技术学院 太原 030006)

**摘 要** 诸多实际应用中,由于确定数据集的类信息通常比较“昂贵”,因此研究者只能为其中很少量的数据标记类信息。针对上述“少量标记数据问题”,文中基于粗糙集理论和信息熵的概念,提出了一种基于信息熵的粗糙特征选择算法。通过分析给定数据集上有标记数据集和无标记数据的信息熵,重新定义了整个数据集上的信息熵。在此基础上定义了半监督意义下基于信息熵的特征重要度,设计了一种基于信息熵的可有效处理含有少量标记数据的半监督粗糙特征选择算法。实验结果进一步验证了所提算法的可行性和高效性。

**关键词** 少量标记数据,信息熵,半监督,特征选择

**中图分类号** TP181 **文献标识码** A

## Semi-supervised Feature Selection Algorithm Based on Information Entropy

WANG Feng LIU Ji-chao WEI Wei

(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

**Abstract** In applications, since it is usually expensive to determine data labels, researchers can only mark a very small amount of data. Hence, on the basis of rough set theory and entropy, this paper proposed an entropy-based rough feature selection algorithm for the problem of “small labeled samples”. In the context of semi-supervised learning, entropy and feature significance were defined in this paper. On this basis, a new semi-supervised feature selection algorithm was proposed to deal with datasets which contain only small labels. Experimental results show that the new algorithm is feasible and efficiency.

**Keywords** Small labeled data, Information entropy, Semi-supervised, Feature selection

## 1 引言

数据挖掘旨在将数据转换为有用信息,即发现数据中的知识,是目前信息化社会中发现知识的重要手段之一。基于数据挖掘技术获取到的知识已在诸多实际应用中被广泛使用,如市场管理、工程决算、企划设计以及投资预测等。对数据挖掘技术的深入探索以及进行的大量研究实例表明,数据预处理对有效使用数据挖掘工具并成功发现知识至关重要<sup>[1-4]</sup>。特征选择是数据挖掘中一种被广泛采用的数据预处理技术,旨在去除高维数据集的冗余特征,保留少量有用特征,从而降低数据存储和转换的成本,节省计算时间,提高数据的紧密度和训练模型的精度<sup>[5-8]</sup>。然而,针对在诸多实际应用领域中获取到的海量高维数据集,由于确定数据的类信息通常会比较“昂贵”,因此只能为其中很少量的数据标记类信息,更多的是“廉价”的不含有类信息的数据,即无类标签的数据。因此,“少量标记数据问题”成为当前特征选择中的一个全新挑战,即如何在有标记数据相对缺乏的情况下有效利用大量的无标记数据进行特征选择,成为了大数据背景下的一个亟待解决的热点研究问题<sup>[9-15]</sup>。

对于数据集中只含有少量有标记数据和大量无标记数据的大数据集,显然适合引入半监督学习的思想来进行有效特

征子集的选取。近年来,半监督特征选择也已经引起了一些研究者的关注,并取得了一定的研究成果<sup>[16-18]</sup>。Zhao 等提出了基于谱分析的半监督特征选择算法<sup>[19]</sup>,该方法是通过归一化的方法对特征进行排序,特征的相关性是通过有标数据和无标数据来估计的。Nakatani 等提出了基于最大密度子图的半监督特征选择算法,该方法利用结合有标记数据和无标记数据共同产生的图结构来进行特征选择<sup>[20]</sup>。Handl 等从多目标优化的角度进行半监督特征的选择,并在实验中验证了在缺乏先验知识时该优化方法可以达到较好的效果<sup>[21]</sup>。Izutani 等基于图理论,首先选取出所有两两相关的特征对构造特征集的邻近图,并通过寻找最大连通子图求解互相关联的特征子集<sup>[22]</sup>。Xu 等基于最大边缘原则和流形正则化概念,提出了一种嵌入式的半监督特征选择方法<sup>[23]</sup>。该方法通过最大边缘原则保证了选取特征的可区分性,流形正则化则保证了由无标记数据构成的流形结构的平滑性。此外,Ren 等提出了将类标记从少量的训练样本扩展到无标记数据上的框架,但并没有充分利用无类标记的信息<sup>[24]</sup>。上述半监督意义下的特征选择算法大都通过有标记数据来对无标记数据进行估计。但如果有标记数据较少,或者只能对原始数据集中的部分无标记数据进行估计时,大量的无标记数据中蕴含的信息就不能被充分利用。

本文受国家自然科学基金项目(61402272,61772323,61603230),山西省教育厅高效科技创新项目(2016111)资助。

王 锋(1984—),女,博士,副教授,CCF 会员,主要研究方向为粒度学习和特征选择,E-mail: sxuwangfeng@126.com;刘吉超(1994—),男,硕士生,主要研究方向为机器学习和特征选择;魏 巍(1980—),男,博士,副教授,主要研究方向为机器学习和粒度计算。

为此,基于信息熵的概念,通过重新定义含有少量标记的数据集上的信息熵,提出一种半监督意义下的粗糙特征选择算法。熵的概念被众多学者引入到信息论中<sup>[25-29]</sup>,用于度量给定数据集中的不确定性,且针对给定的有标记数据集和无标记数据集,均给出了其信息熵的定义<sup>[30-32]</sup>。因此,本文基于互补信息熵的概念,对给定数据集上的有标记数据、无标记数据和通过有标记数据标注的数据分别给出了其不确定度量,并基于上述3部分数据熵的组合,定义了含有少量数据标记数据集上的互补信息熵。在此基础上,重新定义特征重要度,进而设计了一种基于信息熵的半监督粗糙特征选择算法。为验证新算法的可行性和高效性,本文选取了6组UCI数据集进行仿真实验,针对每组数据集,随机选取其中10%作为有标记数据,剩余90%作为无标记数据,使用本文算法和有监督的基于互补信息熵的粗糙特征选择算法进行比较和测试,实验结果进一步证明了新算法的有效性。

本文第2节基于粗糙集理论介绍了含有少量标记数据集的符号表示和互补信息熵的定义;第3节提出了一种基于K-近邻思想的无标记数据标记算法,并定义了半监督意义下的互补信息熵,同时重新定义了特征重要度的度量,并给出了基于信息熵的半监督粗糙特征选择算法;第4节选取了6组UCI数据集进行实验测试和分析;最后总结全文。

## 2 基本概念

给定的数据集  $S$  可表示为一个四元组  $S = (U, A, V, f)$ , 其中  $U$  是数据集的论域,表示所有数据对象的集合;  $A$  是描述数据对象的特征集,即属性集;  $V_a$  表示特征  $a \in A$  的值域,且有  $V = \bigcup_{a \in A} V_a$ ; 对于任意的  $a \in A$  和  $x \in U$ ,  $f: U \times A \rightarrow V$  是一个信息函数,且有  $f(x, a) \in V_a$ 。为方便,上述表示中  $S$  通常可简写为  $S = (U, A)$ 。

特征子集  $B \subseteq A$  可诱导一个不可区分关系,表示为  $R_B = \{(x, y) \in U \times U \mid f(x, a) = f(y, a), \forall a \in B\}$ 。关系  $R_B$  称为等价关系,可将论域划分为一组等价类,表示为  $U/R_B = \{[x]_B \mid x \in U\}$ ,通常简写为  $U/B$ ,其中  $[x]_B = \{y \in U \mid (x, y) \in R_B\}$ 。

给定的有标记数据集  $S$  可表示为  $S = (U, C \cup D)$ ,且  $C \cap D = \emptyset$ 。其中,  $C$  为特征集,  $D$  表示类标记。

**定义1(互补信息熵)**<sup>[22]</sup> 给定无标记数据集  $S = (U, A)$ ,且  $B \subseteq A$ 。假设  $U/B = \{X_1, X_2, \dots, X_m\}$ ,则数据集在  $B$  上的互补信息熵定义为:

$$E(B) = \sum_{i=1}^m \frac{|X_i|}{|U|} \left(1 - \frac{|X_i|}{|U|}\right) \quad (1)$$

**定义2(互补条件熵)**<sup>[22]</sup> 给定有标记数据集  $S = (U, C \cup D)$ ,且  $B \subseteq C$ 。假设  $U/B = \{X_1, X_2, \dots, X_m\}$  和  $U/D = \{Y_1, Y_2, \dots, Y_n\}$ ,则数据集在  $B$  相对于  $D$  上的互补条件熵定义为:

$$E(D|B) = \sum_{i=1}^m \sum_{j=1}^n \frac{|Y_i \cap X_j|}{|U|} \frac{|Y_i \cap X_j^c|}{|U|} \quad (2)$$

对于给定的含有少量标记数据集  $S$ ,其论域  $U$  可表示为  $U = U_l \cup U_u$ ,且  $U_l \cap U_u = \emptyset$ ,其中  $U_l$  表示  $S$  中的有标记数据,  $U_u$  表示  $S$  中的无标记数据。

## 3 半监督意义下的信息熵

基于定义1和定义2中信息熵的概念,本节介绍含有少

量标记数据集上半监督意义下的信息熵。

### 3.1 无标记数据标记算法

考虑到数据集中只含有少量的有标记数据,使用有标记数据对全部的无标记数据进行标记显然不够准确。本文引入K-近邻的思想,即计算距离每个有标记数据最近的  $K$  个近邻,并对其进行标记。对于做了标记的数据,如果某个数据被标记为多于一个的类别,则选择其中距离最近的有标记数据的标签进行标记。通过有标记数据和  $K$ -近邻思想,将做了标记的数据记为  $U_h$ 。具体的标记算法如算法1所示。

**算法1** 基于K-近邻的无标记数据标记算法

输入:含有少量标记的数据集  $S = (U_l \cup U_u, C \cup D)$

输出:标记了类别信息的数据集  $U_h$

Step1  $U_h \leftarrow \emptyset$ ,对于任意的有标记数据  $x_i \in U_l (i=1, 2, \dots, |U_l|)$ ,计算距离  $x_i$  最近的  $k$  个近邻。

Step2 将 Step1 中计算得到的距离  $x_i$  最近的  $k$  个近邻  $\{x_1^i, x_2^i, \dots, x_k^i\}$  标记为与  $x_i$  相同的类标签,且  $U_h \leftarrow U_h \cup \{x_1^i, x_2^i, \dots, x_k^i\}$ 。

Step3 如果存在  $x \in U_h$  被标记了多于一个的类标签,那么选择距离  $x$  最近的有标记数据的类标签作标记。

Step4 算法结束,返回  $U_h$ 。

基于上述标记算法,可对给定数据集中的部分无标记数据进行标记。K-近邻思想的引入是基于数据集中相邻数据对象属于同一类别的假设的。但考虑到对所有无标记数据进行标注比较耗时,而且会降低类信息标注的准确性,本文基于K-近邻的思想,只对无标记数据中的部分数据进行标注。

### 3.2 半监督意义下的信息熵

对于给定的含有少量标记的数据集  $S$ ,使用信息熵度量数据集  $S$  中无标记数据集的不确定性,使用条件熵度量有标记数据集的不确定性。令  $E_u$  表示无标记数据  $U_u$  上的信息熵,  $E_l$  表示有标记数据  $U_l$  上的条件熵,令  $E_h$  表示  $U_h$  上的条件熵。  $E_u$ ,  $E_l$  和  $E_h$  的具体定义如下。

**定义3** 对于给定的含有少量标记的数据集  $S = (U_l \cup U_u, C \cup D)$ ,且  $B \subseteq C$ ,假设  $U_l/B = \{X_1^l, X_2^l, \dots, X_m^l\}$ ,  $U_u/B = \{X_1^u, X_2^u, \dots, X_n^u\}$ ,  $U_h/B = \{X_1^h, X_2^h, \dots, X_k^h\}$  和  $U/D = \{Y_1, Y_2, \dots, Y_n\}$ ,其中  $U_l/B$ ,  $U_u/B$  和  $U_h/B$  分别表示  $U_l$ ,  $U_u$  和  $U_h$  在  $B$  上的划分,则数据集  $S$  中的互补熵  $E_u$ ,  $E_l$  和  $E_h$  分别定义为:

$$E_u(B) = \sum_{i=1}^n \frac{|X_i^u|}{|U_u|} \left(1 - \frac{|X_i^u|}{|U_u|}\right) \quad (3)$$

$$E_l(D|B) = \sum_{i=1}^m \sum_{j=1}^n \frac{|Y_j \cap X_i^l|}{|U_l|} \frac{|Y_j \cap X_i^{lc}|}{|U_l|} \quad (4)$$

$$E_h(D|B) = \frac{E_h(B) - E_h(D|B)}{E_h(B)} \quad (5)$$

**定义4** 对于给定的含有少量标记的数据集  $S = (U_l \cup U_u, C \cup D)$ ,且  $B \subseteq C$ ,则半监督意义下数据集  $S$  的信息熵定义为:

$$E_r(B|D) = \alpha E_l + \beta E_u + \frac{|U_h|}{|U|} \cdot E_h \quad (6)$$

其中,  $\alpha, \beta$  的值在本文中分别取 0.9 和 0.1。这是由于信息熵  $E_l$  在由特征集诱导的协调数据集上的值都为 0,当数据集中存在不协调数据时,  $E_l$  的值才会大于 0。而  $E_u$  的值通常会大于 0,只有当数据集的划分是全集(即数据取值全部相同)时,才会等于 0。因此为平衡  $E_l$  和  $E_u$  的权重,本文将  $E_l$  和  $E_u$  的权值定为 0.9 和 0.1。

### 3.3 特征重要度

基于上述半监督意义下信息熵的概念,定义特征重要度如下。

**定义 5** 对于给定的含有少量标记的数据集  $S = (U_l \cup U_u, C \cup D)$ , 且  $B \subseteq C$ , 任意特征  $a \in B$  的特征重要度定义为:

$$\text{sig}_m(a, B, D) = E_r(D|B - \{a\}) - E_r(D|B) \quad (7)$$

**定义 6** 对于给定的含有少量标记数据集  $S = (U_l \cup U_u, C \cup D)$ , 且  $B \subseteq C$ , 任意特征  $a \in C - B$  的属性重要度定义为:

$$\text{sig}_{out}(a, B, D) = E_r(D|B) - E_r(D|B \cup \{a\}) \quad (8)$$

粗糙集理论中,定义 5 中的特征重要度通常称为内部重要度,主要用于删除特征子集中的冗余特征,如果某个特征的内部重要度为零,则表示该特征不能更好地提供有效的信息,即是冗余的。定义 6 中的特征重要度通常被称为外部重要度,主要用于向当前特征子集中添加新的特征。

### 3.4 半监督特征选择算法

**算法 2** 基于信息熵的半监督特征选择算法(SFSE)

输入:含有少量标记的数据集  $S = (U, C \cup D)$ ,  $U_u$  和  $U_l$

输出:特征选择结果 R

Step1 随机选取  $a \in C, B \leftarrow a$ ;

Step2 while  $E_r(B|D) \neq E_r(C|D)$

do{对于任意特征  $a \in C - B$ , 计算特征重要度  $\text{sig}_{out}(a, B, D)$ , 选择外部重要度最大的特征  $a_0, a_0 = \max\{\text{sig}_{out}(a, B, D)\}, a \in C - B$ ;  
 $B \leftarrow B \cup \{a_0\}$ ;  
}

Step3 while  $|E_r(D|B) - E_r(D|C)| \geq \epsilon$

do{对于任意特征  $a \in B$ , 计算特征重要度  $\text{sig}_m(a, B, D)$ ;  
如果  $\text{sig}_m(a, B, D) = 0$ , 则  $B \leftarrow B - \{a\}$ ;  
}

Step4  $R \leftarrow B$ , 算法结束, 返回 R。

上述算法 Step3 中的  $\epsilon$  是用于控制算法搜索结束的阈值,本文中取  $\epsilon = 0.0001$ 。由于本文中只是重新定义了特征的

重要度,而上述算法的搜索过程和基于信息熵的有监督特征选择<sup>[7,25]</sup>的搜索过程是相同的,因此本文未对上述算法的时间复杂度作分析。

## 4 实验分析

为验证本文算法 SFSE 的有效性,选取 6 组 UCI 数据集(见表 1)进行仿真实验。由于互补信息熵(定义 1)和互补条件熵(定义 2)均是基于符号数据定义的,因此本节实验选取的数据集为符号数据集。针对符号数据求解算法 1 中近邻的距离公式如下:

$$d(x_1, x_2) = \sum_{i=1, a_i \in \{C\}}^{|C|} \delta(f(x_1, a_i), f(x_2, a_i))$$

其中,  $x_1, x_2 \in U, \delta(x, y) = \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases}$ 。程序运行的个人计算机配置为 CPU Inter(R) Core(TM) i7-6700, 3.40 GHz, 内存为 8.00 GB, 操作系统是 Windows 7。程序开发平台是 Microsoft Visual Studio 2005, 使用 C# 编程语言。实验中对于

表 1 中的每组数据集,随机抽取其中 10% 作为有标记数据,剩余的 90% 作为无标记数据。利用本文提出的半监督特征选择算法进行有效特征的选取,并与基于互补信息熵的有监督的特征选择算法 IFS<sup>[7,28]</sup>进行特征选择结果的比较。针对特征选择结果,实验中选取了机器学习中两个常见的分类器(即朴素贝叶斯(NBC)和决策树(C4.5))来比较特征选择结果的分类精度。实验结果如表 2 所列。

表 1 数据集的描述

| 数据集         | 样本数   | 特征数 | 类别数 |      |
|-------------|-------|-----|-----|------|
|             |       |     | NBC | C4.5 |
| Cancer      | 683   | 9   | 2   |      |
| Tic-tac-toe | 958   | 9   | 2   |      |
| Kr-vs-kp    | 3196  | 36  | 2   |      |
| Mushroom    | 5644  | 22  | 2   |      |
| Letter      | 20000 | 16  | 26  |      |
| Krkopt      | 28056 | 6   | 18  |      |

表 2 特征选择结果及分类精度的比较

| 数据集         | 算法   | 特征选择结果  | 分类精度            |                 |
|-------------|------|---|-----------------|-----------------|
|             |      |   | NBC             | C4.5            |
| Cancer      | SFSE | 1,2,7,9   | 0.9604 ± 0.0470 | 0.9253 ± 0.0956 |
|             | IFS  | 1,3,5,6   | 0.9677 ± 0.0351 | 0.9209 ± 0.0954 |
| Tic-tac-toe | SFSE | 1,3,4,5,9   | 0.7035 ± 0.3786 | 0.7662 ± 0.3248 |
|             | IFS  | 1,2,3,4,5,6,7,9   | 0.7212 ± 0.3713 | 0.8382 ± 0.1899 |
| Kr-vs-kp    | SFSE | 1,2,3,4,5,7,8,9,11,12,14,15,16,17,19,20,21,22,23,25,27,28,29,30,31,32,35        | 0.7274 ± 0.3190 | 0.7978 ± 0.2529 |
|             | IFS  | 1,3,4,5,6,7,9,10,11,12,13,15,16,17,18,20,21,23,24,25,26,27,28,30,31,33,34,35,36 | 0.8826 ± 0.2225 | 0.9909 ± 0.0128 |
| Mushroom    | SFSE | 1,3,5,7,9   | 0.9848 ± 0.0224 | 0.9942 ± 0.0092 |
|             | IFS  | 1,3,5,20  | 0.9895 ± 0.0245 | 0.9970 ± 0.0053 |
| Letter      | SFSE | 2,7,8,9,10,11,12,13,14,15,16  | 0.6530 ± 0.0331 | 0.8832 ± 0.0105 |
|             | IFS  | 1,2,4,5,8,9,10,11,12,13,15,16   | 0.6025 ± 0.0381 | 0.8751 ± 0.0111 |
| Krkopt      | SFSE | 1,2,3,4,5,6   | 0.3605 ± 0.0867 | 0.5658 ± 0.0583 |
|             | IFS  | 1,2,3,4,5,6   | 0.3605 ± 0.0867 | 0.5658 ± 0.0583 |

表 2 分别给出了两种算法在 6 组数据集上的特征选择结果及其在两种分类器下的分类精度。该实验结果中,在数据集 Kr-vs-kp 下由半监督特征选择算法 SFSE 计算得到的特征选择结果的分类精度略低于由算法 IFS 计算得到的精度。在其余的 5 组数据集上,两种算法计算得到的特征选择结果的分类精度都比较接近,没有明显的优或劣。因此,实验结果表明,本文提出的基于信息熵的半监督特征选择算法 SFSE 可找到一个与基于信息熵的有监督的特征选择算法分类性能相近的特征子集。另外,由于算法 SFSE 是半监督处理机制,只

需对给定数据集中的部分数据做类信息标注,即可进行特征子集的选取。综上,对于给定的少量标记的数据集,本文提出的新算法 SFSE 通过对部分无标记数据进行标注后,即可进行有效特征子集的选取,有效降低了数据获取中的标注代价,提高了信息获取和知识发现的效率。

**结束语** 大数据时代的到来为传统的针对“少量标记数据问题”的半监督处理机制带来了全新的挑战,如何高效且快捷地获取信息迅速成为一个亟待解决的科学问题。本文从降低标注代价角度,结合粗糙集理论,提出了一种基于信息熵的

半监督特征选择方法。实验结果进一步表明,本算法可以找到一个有效的特征子集,而且有效降低了对无标记数据的标注代价。后期的研究工作将继续深入探索更有效的半监督处理机制下的高效特征选择方法。

### 参考文献

- [1] BLUM A L, LANGLEY P. Selection of relevant features and examples in machine learning [J]. *Artificial Intelligence*, 1997, 97(1-2): 245-271.
- [2] DASH M, CHOI K, SCHEUERMANN P, et al. Feature selection for clustering—a filter solution[C]// *Proceedings of the Second International Conference on Data Mining*, 2002: 115-122.
- [3] LIU H, YU L. Toward integrating feature selection algorithms for classification and clustering [J]. *IEEE Transaction on Knowledge and Data Engineering*, 2005, 17(4): 491-502.
- [4] HU Q H, YU D R, LIU J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. *Information Sciences*, 2008, 178(18): 3577-3594.
- [5] HU Q H, XIE Z X, YU D R. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation [J]. *Pattern Recognition*, 2007, 40(12): 3509-3521.
- [6] CHEN H M, LI T R, QIAO S J, et al. A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values [J]. *International Journal of Intelligent Systems*, 2010, 25(10): 1005-1026.
- [7] LIANG J Y, WANG F, DANG C Y, et al. An efficient rough feature selection algorithm with a multi-granulation view [J]. *International Journal of Approximate Reasoning*, 2012, 53(6): 912-926.
- [8] LIU D, LI T R, RUAN D, et al. An incremental approach for inducing knowledge from dynamic information systems [J]. *Fundamenta Informaticae*, 2009, 94(2): 245-260.
- [9] LI T R, RUAN D, GEERT W, et al. A rough sets based characteristic relation approach for dynamic attribute Fgeneralization in data mining [J]. *Knowledge-Based Systems*, 2007, 20(5): 485-494.
- [10] JING Y G, LI T R, HUANG J F, et al. An incremental attribute reduction approach based on knowledge granularity under the attribute generalization [J]. *International Journal of Approximate Reasoning*, 2016, 76: 80-95.
- [11] JING Y G, LI T R, FUJITA H, et al. An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view [J]. *Information Sciences*, 2017, 411: 23-38.
- [12] JING Y G, LI T R, HUANG J F, et al. A group incremental reduction algorithm with varying data values [J]. *International Journal of Intelligent Systems*, 2016, 32(9): 900-925.
- [13] 杨明. 一种基于改进差别矩阵的核增量式更新算法[J]. *计算机学报*, 2006, 29(3): 407-413.
- [14] 黄兵, 周献中, 张蓉蓉. 基于信息量的不完备信息系统属性约简[J]. *系统工程理论与实践*, 2005, 4(4): 55-60.
- [15] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为  $\max(O(|C||U|), O(|C|^2|U/C|))$  的快速属性约简算法[J]. *计算机学报*, 2006, 29(3): 391-399.
- [16] ZHAO M Y, JIAO L C, MA W P, et al. Classification and saliency detection by semi-supervised low-rank representation [J]. *Pattern Recognition*, 2016, 51(C): 281-294.
- [17] BENABDESLEM K, HINDAWI M. Efficient Semi-supervised Feature Selection: Constraint, Relevance and Redundancy [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(5): 1131-1143.
- [18] FORESTIER G, WEMMERT C. Semi-supervised learning using multiple clustering with limited labeled data [J]. *Information Sciences*, 2016, 361-362(C): 48-65.
- [19] ZHAO Z, LIU H. Semi-supervised feature selection via spectral analysis[C]// *SIAM International Conference on Data Mining (SDM 2007)*, 2007.
- [20] NAKATANI Y, ZHU K, UEHARA K. Semi-supervised learning using feature selection based on maximum density subgraphs [J]. *Systems and Computers in Japan*, 2007, 38(9): 32-43.
- [21] HANDL J, KNOWLES J. Semi-supervised feature selection via multi-objective optimization[C]// *The 2006 International Joint Conference on Neural Networks*, 2006.
- [22] IZUTANI A, UEHARA K. A Modeling Approach Using Multiple Graphs for Semi-Supervised Learning [C]// *International Conference on Discovery Science*. Springer-Verlag, 2008: 296-307.
- [23] XU Z L, KING I, MICHAEL R-T L, et al. Discriminative Semi-Supervised Feature Selection Via Manifold Regularization [J]. *IEEE Transactions on Neural Networks*, 2010, 21(7): 1033-1046.
- [24] REN J T, QIU Z Y, FAN W, et al. Forward semi-supervised feature selection[C]// *Proceedings of the 12th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining (PAKDD'08)*. Berlin: Springer-Verlag, 2008: 970-976.
- [25] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. *计算机学报*, 2002, 25(7): 759-766.
- [26] LIANG J Y, CHIN K S, DANG C Y, et al. A new method for measuring uncertainty and fuzziness in rough set theory [J]. *International Journal of General Systems*, 2002, 31(4): 331-342.
- [27] LIANG J Y, SHI Z Z, LI D Y, et al. The information entropy, rough entropy and knowledge granulation in incomplete information systems [J]. *International Journal of General Systems*, 2006, 34(1): 641-654.
- [28] WANG F, LIANG J Y, QIAN Y H. Attribute reduction: a dimension incremental strategy [J]. *Knowledge-Based Systems*, 2013, 39(2): 95-108.
- [29] LIANG J Y, WANG F, DANG C Y, et al. A group incremental approach to feature selection applying rough set technique [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(2): 294-308.
- [30] 王锋, 魏巍. 缺失数据数据集的组增量式特征选择[J]. *计算机科学*, 2015, 42(7): 285-290.
- [31] 刘薇, 梁吉业, 魏巍, 等. 一种基于条件熵的增量式属性约简求解算法[J]. *计算机科学*, 2011, 38(1): 229-231, 239.
- [32] QIAN Y H, LIANG J Y, PEDRYCZ W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory [J]. *Artificial Intelligence*, 2010, 174(9-10): 597-618.