

基于灰关系分析的特征重要性评估及应用

生佳根, 胡雨晴

(江苏科技大学 计算机科学与工程学院, 江苏 镇江 212003)

摘 要: 灰关系分析因为能够度量参考样本和比较样本间的相似性而被广泛应用于聚类和分类算法, 尤其是样本信息不完全以及样本量较小时. 其时间域通常为横向的每个样本的各个特征的数据. 本文则从一个不同的角度, 将灰关系的时间域变化为垂直方向的每个特征下各个样本的数据, 并将排序问题的结果或者分类问题的标号作为其参考向量, 通过特征对应列和参考向量之间的灰关系分析进而度量特征的重要性, 从而为基于重要性的特征选择提供决策依据. 该方法可广泛应用于大量特征但样本稀少情形下的降维, 并能提供解释性良好的特征重要性参数. 为验证这种重要性评估方法, 在田径十项全能项目以及 IRIS 数据集上进行了基于灰关系的特征重要性评估实验, 验证了该方法的实用价值.

关键词: 灰关系分析; 相似性; 模糊聚类; 核机理论

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-7180(2012)11-0166-06

Grey Relational Analysis Based Feature Importance Evaluation and Its Application

SHENG Jia-gen, HU Yu-qing

(School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract: Since Grey Relational Analysis (GRA) can measure the similarity between the reference samples and the compared samples, it is widely applied in clustering and classification, especially, on the condition that the information of samples is incomplete and the size of samples is small. In general, the time domain of gray relation is the data of the horizontal features of each sample. Differed from the traditional GRA, a novel aspect is revealed, where the time domain of gray relation becomes the data of the vertical features of each sample. After defining the reference vector by ranking or labels in classification, grey relational analysis is conducted between the corresponding column of each feature and the referenced vector, thus, the feature importance is computed by the grey relational grade. Then feature selection can be performed with the feature importance, so as to conduct dimension reduction for sparse and large number of features. Worthy to be pointed is that the presented method can provide the parameters of features importance with good interpretation ability. To verify the importance evaluation method, experiments are performed on Decathlon and IRIS data set and the experimental results show that it is consistent to the prior knowledge.

Key words: grey relational analysis (GRA); feature importance; ranking; evaluation

1 引言

在分类和聚类问题中采集的数据中经常面对数目繁多的各种特征, 对这些特征的重要性一般都简化为相等, 然而这些特征之间和分类或者聚类的结

果之间存在贡献不等的现象, 因而揭示特征的重要性具有很好的理论和实用价值. 并且和一般的特征抽取方法(如主成分分析)缺乏完整的物理涵义不同, 特征重要性的评估不损失特征的物理涵义, 因而可显著提高聚类或分类的物理解释能力, 并可进一

步指导并应用于特征选择中,为数据的降维提供一种可解释的采样方法^[1]. Yang 等^[2]通过计算使用和不使用每个特征的信息增益进行特征重要性计算, Galavotti 等^[3]通过计算每个特征的差方统计量评估特征的重要性, Dash 等^[4]通过类标号和特征之间的相关性计算进行重要性排序,而 Liu 等^[1]在回顾目前的文本表示的向量空间构成成分基础上提出了词贡献量,其在 td-idf 基础上,用所有文档之间的同一词的 td-idf 之积进行累加定义词贡献量.但是上述的这些方法基于的统计学习基础,一般都基于充分统计的基础上,且计算量大等不足.

灰色理论^[5]是一种专门研究信息不完全,数据有限的系统的行为理论,其广泛应用于模式分类、聚类以及故障检测等领域.灰关系分析^[5]是灰色理论中的一个重要分支,其刻画了参考样本和比较样本之间的相似性.其首先对数据进行转换,并计算灰关系系数矩阵,对各个特征的灰关系系数求和获得灰色关联度,从而获得各样本和参考样本的灰色关联度的排序.因其计算简单,且适合稀疏小样本数据集,广泛应用与环境污染评估^[6]、经销商评估^[7]、体育十项全能^[8]的评估等场景中.然而这些应用中,都设定某级别的标准特征值,然后将实际样本作为参考向量,和前面的标准特征值进行比对,计算实际样本和某一级别的灰色关联度最大的作为评判.本文利用灰关系的基础理论,将原先的参考量定义为和特征并列的样本标号,计算每个特征维度和样本标号维度的相似性,并计算每个特征的灰色关联度,则和参考量具有最高灰色关联度的特征具有最高的重要性,反之则不然,从而通过简单的计算对特征的重要性进行评估.

2 灰关系分析

灰关系分析^[1-9]针对参考样本和比较样本之间相似性进行研究,假设 p 维参考样本集 x_r 和比较样本集 $X_c = \{x_j\}_{j=1}^m (m \geq 1)$, 则参考样本 x_r 和比较样本 x_j 之间的第 k 维灰关系系数是:

$$g(x_r(k), x_j(k)) = \frac{\Delta_{\min} + \xi \cdot \Delta_{\max}}{\Delta_j(k) + \xi \cdot \Delta_{\max}} \quad (1)$$

其中 $\Delta_j(k) = |x_r(k) - x_j(k)|$, $\Delta_{\min} = \min_j \min_k \{\Delta_j(k)\}$, $\Delta_{\max} = \max_j \max_k \{\Delta_j(k)\}$, ξ 是一个 0 到 1 之间的分辨率控制参数.每一维的灰关系系数计算

后, x_r 和 x_j 间的灰色关联度可采用每一维加权平均的方法计算:

$$\begin{aligned} \lambda_j &= \lambda(x_r, x_j) \\ &= \sum_{k=1}^p \alpha_k g(x_r(k), x_j(k)) \\ \text{s. t. } \sum_{k=1}^p \alpha_k &= 1 \end{aligned} \quad (2)$$

从而可构建参考样本和比较样本集之间的灰关系关联度.由(1)和(2)可知灰关系系数和灰色关联度都在 0 和 1 之间,且两个样本之间差异越大,灰关系等级越小,反之亦然.

基于拓扑的灰关系分析^[1, 9]中,对样本 x_i 归一化为 x_i^* 有三种方法:

(1)“越小越好”,例如运动员跑步的时间,计算

$$x_i^*(k) = \frac{\max_l x_l(k) - x_i(k)}{\max_l x_l(k) - \min_l x_l(k)};$$

(2)“越大越好”,如运动员跳远的成绩,计算

$$x_i^*(k) = \frac{x_i(k) - \min_l(x_l(k))}{\max_l x_l(k) - \min_l x_l(k)};$$

(3)“中间最好”,如运动员的年龄,计算

$$x_i^*(k) = 1 - \frac{x_i(k) - x_o(k)}{\max\{\max_l x_l(k) - x_o(k), x_o(k) - \min_l x_l(k)\}}$$

其中 $x_o(k)$ 为任一指定值.然后计算 x_n^* 和 x_{rj}^* 的灰关系系数为:

$$\begin{aligned} \lambda(x_n, x_{rj}) &= \lambda(x_n^*, x_{rj}^*) \\ &= \sum_{k=1}^p \alpha_k g(x_n^*(k), x_{rj}^*(k)) \end{aligned} \quad (3)$$

如无特殊设置,一般各维特征的 α_k 等权处理.

3 基于灰关系分析的重要性评估方法

下面以 2011 年世界田径锦标赛十项全能(Decathlon)的成绩为例,进行特征重要性评估.剔除参赛项目不全或者无成绩的选手,余下 21 个选手,其各项成绩及比赛得分如表 1 所示.

由于 100 米、400 米、110 米跨栏以及 1500 米赛跑成绩越小越好,可根据最小最好准则进行归一化,其余的采用最大最好准则归一化后得到的新成绩如表 2 所示.

表 1 2011 年世界田径锦标赛成绩

序号	100 米	400 米	跨栏	1500 米	跳远	铅球	跳高	铁饼	撑杆跳	标枪	得分
1	10.55	48.37	13.97	285.7	7.45	15.09	2.02	49.89	4.8	68.99	8607
2	10.46	46.99	13.85	258.9	7.46	14.44	2.02	46.17	4.6	55.17	8505
3	11.07	49.17	14.29	264.2	7.33	14.54	2.05	46.25	5	69.12	8501
4	11.34	51.35	15.49	281.7	7.45	16.17	2.14	50.29	5	64.8	8313
5	10.76	48.35	14.42	265.4	7.29	14.13	1.93	42.23	5.2	61.07	8298
6	10.81	47.73	14.89	266.1	7.41	13.76	2.02	43.97	4.9	58.93	8256
7	11.08	49.9	14.33	276.6	6.8	16.01	1.93	48.56	4.9	66.5	8211
8	11.18	49.46	14.92	283.1	7.3	16.06	1.96	47.93	4.7	68.42	8200
9	11.28	50.65	14.54	275.4	7.12	14.76	2.02	47.16	4.9	66.4	8164
10	10.88	47.34	14.56	255	7.42	13.11	1.96	37.84	4.9	59	8158
11	11.2	49.5	14.37	269.2	7.06	14.92	1.99	41.65	5	63.25	8134
12	10.75	48.46	14.65	269.4	7.59	14.43	1.99	43.74	4.7	52.16	8132
13	11.2	49.46	14.79	285.9	7.79	12.76	2.17	37.2	5.1	58.91	8069
14	11.25	51.18	14.75	296.5	7.3	15.2	2.05	46.93	4.8	67.28	8069
15	10.85	49.47	14.04	292.3	7.21	14.61	1.99	43.39	4.7	57.35	7982
16	10.71	48.48	14.25	287.3	6.74	14.77	1.9	46.46	4.7	54.38	7902
17	11.11	49.24	14.95	255.6	7.24	12.96	1.96	39.53	4.9	53.33	7860
18	10.56	50.12	14.19	294.4	7.35	13.26	2.02	41.71	4.5	54.61	7832
19	11	49.95	14.86	287.3	7.31	13.75	1.96	45.65	4.5	55.69	7761
20	11.42	50.89	15.2	283.9	6.96	12.88	2.02	43.84	4.4	67.73	7639
21	11.24	52.01	14.64	298.4	6.86	15.69	1.93	47.1	4.8	46.91	7550

表 2 归一化后的 2011 年世界田径锦标赛成绩

序号	100 米	400 米	100 米跨栏	1500 米	跳远	铅球	跳高	铁饼	撑杆跳	标枪
1	0.91	0.73	0.93	0.29	0.68	0.68	0.44	0.97	0.50	0.99
2	1.00	1.00	1.00	0.91	0.69	0.49	0.44	0.69	0.25	0.37
3	0.36	0.57	0.73	0.79	0.56	0.52	0.56	0.69	0.75	1.00
4	0.08	0.13	0.00	0.38	0.68	1.00	0.89	1.00	0.75	0.81
5	0.69	0.73	0.65	0.76	0.52	0.40	0.11	0.38	1.00	0.64
6	0.64	0.85	0.37	0.74	0.64	0.29	0.44	0.52	0.63	0.54
7	0.35	0.42	0.71	0.50	0.06	0.95	0.11	0.87	0.63	0.88
8	0.25	0.51	0.35	0.35	0.53	0.97	0.22	0.82	0.38	0.97
9	0.15	0.27	0.58	0.53	0.36	0.59	0.44	0.76	0.63	0.88
10	0.56	0.93	0.57	1.00	0.65	0.10	0.22	0.05	0.63	0.54
11	0.23	0.50	0.68	0.67	0.30	0.63	0.33	0.34	0.75	0.74
12	0.70	0.71	0.51	0.67	0.81	0.49	0.33	0.50	0.38	0.24
13	0.23	0.51	0.43	0.29	1.00	0.00	1.00	0.00	0.87	0.54
14	0.18	0.17	0.45	0.04	0.53	0.72	0.56	0.74	0.50	0.92
15	0.59	0.51	0.88	0.14	0.45	0.54	0.33	0.47	0.38	0.47
16	0.74	0.70	0.76	0.26	0.00	0.59	0.00	0.71	0.38	0.34
17	0.32	0.55	0.33	0.99	0.48	0.06	0.22	0.18	0.63	0.29
18	0.90	0.38	0.79	0.09	0.58	0.15	0.44	0.34	0.13	0.35
19	0.44	0.41	0.38	0.26	0.54	0.29	0.22	0.65	0.13	0.40
20	0.00	0.22	0.18	0.33	0.21	0.04	0.44	0.51	0.00	0.94
21	0.19	0.00	0.52	0.00	0.11	0.86	0.11	0.76	0.50	0.00

和一般的灰关系分析不同,本方法中采用的参考向量为转换后的比赛得分向量:
$$x_r=[1\ 0.9\ 0.9\ 0.72\ 0.71\ 0.67\ 0.63\ 0.61\ 0.58\ 0.58\ 0.55\ 0.55\ 0.49\ 0.49\ 0.41\ 0.33\ 0.29\ 0.27\ 0.20\ 0.08\ 0]^T$$

为了折中 $\Delta_{\min}$ 和 $\Delta_{\max}$,去(1)中 $\xi=0.5$,然后根据(1)计算灰关系系数矩阵,如表 3 所示.
将 21 个运动员等权处理,根据灰色关联度的计算公式,可得到 10 个特征的灰色关联度,如表 4 所示.

表 3 灰关系系数矩阵

序号	100 米	400 米	跨栏	1500 米	跳远	铅球	跳高	铁饼	撑杆跳	标枪
1	0.09	0.27	0.07	0.71	0.32	0.32	0.56	0.03	0.50	0.01
2	0.10	0.10	0.10	0.01	0.22	0.41	0.46	0.22	0.65	0.53
3	0.54	0.33	0.17	0.11	0.34	0.38	0.34	0.21	0.15	0.10
4	0.64	0.59	0.72	0.34	0.05	0.28	0.17	0.28	0.03	0.08
5	0.02	0.02	0.06	0.05	0.18	0.31	0.60	0.32	0.29	0.07
6	0.03	0.18	0.30	0.08	0.03	0.37	0.22	0.15	0.04	0.13
7	0.27	0.21	0.08	0.12	0.57	0.33	0.51	0.24	0.00	0.26
8	0.36	0.11	0.27	0.26	0.08	0.35	0.39	0.20	0.24	0.35
9	0.44	0.31	0.00	0.05	0.22	0.01	0.14	0.18	0.04	0.30
10	0.01	0.36	0.01	0.42	0.07	0.47	0.35	0.53	0.05	0.03
11	0.32	0.05	0.13	0.12	0.25	0.08	0.22	0.21	0.20	0.18
12	0.15	0.16	0.04	0.12	0.26	0.06	0.22	0.05	0.18	0.31
13	0.26	0.02	0.06	0.20	0.51	0.49	0.51	0.49	0.38	0.05
14	0.31	0.33	0.04	0.45	0.04	0.22	0.06	0.25	0.01	0.43
15	0.19	0.10	0.48	0.27	0.04	0.13	0.08	0.06	0.03	0.06
16	0.41	0.37	0.42	0.08	0.33	0.26	0.33	0.37	0.04	0.00
17	0.03	0.26	0.04	0.69	0.18	0.23	0.07	0.12	0.33	0.00
18	0.63	0.11	0.53	0.17	0.31	0.12	0.18	0.08	0.14	0.08
19	0.24	0.21	0.18	0.06	0.34	0.09	0.02	0.45	0.07	0.20
20	0.08	0.14	0.09	0.25	0.13	0.05	0.36	0.42	0.08	0.85
21	0.19	0.00	0.52	0.00	0.11	0.86	0.11	0.76	0.50	0.00

表 4 灰色关联度和特征重要性排序

特征名	100 米	400 米	100 米跨栏	1500 米	跳远	铅球	跳高	铁饼	撑杆跳	标枪
灰色关联度	14.24	15.02	15.47	15.06	14.63	13.64	13.48	13.71	15.69	15.81
排序	4	6	8	7	5	2	1	3	9	10

下面简要分析该排序的合理性,根据表 1,第二名的成绩其十项成绩分别排名如下:第一、第一、第一、第三、第三、第十一、第五、第九、第十七、第六.而第一名十项成绩排名分别是:第二、第五、第二、第七、第四、第六、第五、第二、第十一、第二.
本应第二名的成绩比第一名要好,但是由于其在关键的铅球、跳高以及铁饼上不如第一名,从而使得其积分落后不少.因而目前十项全能运动员整体水平基础上运动员应重点加强跳高、铅球及铁饼项

目的训练.

4 基于灰关系分析的重要性评估方法的分类应用

4.1 IRIS 数据集

IRIS 数据集是 UCI^[10]提供的 3 种花的数据,其测量了三种花的花萼和花瓣的长度和宽度,其中每种花的数据各 50 组,其统计特性如表 5 所示.
根据上述结果,花瓣长度(petal length)和花瓣

宽度(petal width)对分类具有强相关性,花萼长度(sepal length)具有相关性,而花萼宽度和分类关系不大,因而可知特征的重要性是花瓣长度和宽度的重要性比花萼长度大,最小的是花萼宽度。

表 5 IRIS 数据集统计特征描述

特征	最小值	最大值	均值	方差	分类相关性
sepal length	4.3	7.9	5.84	0.83	0.7826
sepal width	2.0	4.4	3.05	0.43	-0.4194
petal length	1.0	6.9	3.76	1.76	0.9490
petal width	0.1	2.5	1.20	0.76	0.9565

4.2 监督型的基于灰关系分析的特征重要性方法

分类问题和和前面的十项全能成绩排序问题不同,其参考向量为样本的标号,通过计算各个特征向量(样本变化)和样本标号向量之间的相似性进行灰关系系数的计算,并进而计算灰色关联度。

为了缩减篇幅,本文从 IRIS 数据集中每类中抽取 10 个样本,构成的数据样本集合如表 6 所示。

表 6 原始数据

花萼长度	花萼宽度	花瓣长度	花瓣宽度	标号	花萼长度	花萼宽度	花瓣长度	花瓣宽度	标号
5.1	3.5	1.4	0.2	1	5.7	2.8	4.5	1.3	2
4.9	3	1.4	0.2	1	6.3	3.3	4.7	1.6	2
4.7	3.2	1.3	0.2	1	4.9	2.4	3.3	1	2
4.6	3.1	1.5	0.2	1	6.6	2.9	4.6	1.3	2
5	3.6	1.4	0.2	1	5.2	2.7	3.9	1.4	2
5.4	3.9	1.7	0.4	1	6.3	3.3	6	2.5	3
4.6	3.4	1.4	0.3	1	5.8	2.7	5.1	1.9	3
5	3.4	1.5	0.2	1	7.1	3	5.9	2.1	3
4.4	2.9	1.4	0.2	1	6.3	2.9	5.6	1.8	3
4.9	3.1	1.5	0.1	1	6.5	3	5.8	2.2	3
7	3.2	4.7	1.4	2	7.6	3	6.6	2.1	3
6.4	3.2	4.5	1.5	2	4.9	2.5	4.5	1.7	3
6.9	3.1	4.9	1.5	2	7.3	2.9	6.3	1.8	3
5.5	2.3	4	1.3	2	6.7	2.5	5.8	1.8	3
6.5	2.8	4.6	1.5	2	7.2	3.6	6.1	2.5	3

经归一化的数据如表 7 所示。将标号向量也归一化作为参考向量,根据灰关系系数计算得到表 8 所示的灰关系系数矩阵。

进而可将每列累加得到每个特征相对于标号向量的灰色关联度[15.64 19.08 14.10 14.62],该结

果和数据集中的统计分析也是一致的。

表 7 归一化后的数据

花萼长度 变换	花萼宽度 变换	花瓣长度 变换	花瓣宽度 变换	花萼长度 变换	花萼宽度 变换	花瓣长度 变换	花瓣宽度 变换
0.7813	0.25	0.9811	0.9583	0.5938	0.6875	0.3963	0.5
0.8438	0.5625	0.9811	0.9583	0.4063	0.3750	0.3585	0.3750
0.9063	0.4375	1	0.9583	0.8438	0.9375	0.6226	0.6250
0.9375	0.5	0.9623	0.9583	0.3125	0.6250	0.3774	0.5
0.8125	0.1875	0.9811	0.9583	0.7500	0.7500	0.5094	0.4583
0.6875	0	0.9245	0.8750	0.4063	0.3750	0.1132	0
0.9375	0.3125	0.9811	0.9167	0.5625	0.7500	0.2830	0.25
0.8125	0.3125	0.9623	0.9583	0.1563	0.5625	0.1320	0.1667
1	0.6250	0.9811	0.9583	0.4063	0.6250	0.18877	0.2917
0.8438	0.5	0.9623	1	0.3438	0.5625	0.1509	0.1250
0.1875	0.4375	0.3585	0.4583	0	0.56250	0	0.1667
0.3750	0.4375	0.3962	0.4167	0.8438	0.8750	0.3962	0.3333
0.2188	0.5	0.3208	0.4167	0.0938	0.6250	0.0566	0.2917
0.6563	1	0.4906	0.5	0.2813	0.8750	0.1509	0.2917
0.3438	0.6875	0.3774	0.4167	0.1250	0.1875	0.0943	0

表 8 灰关系系数矩阵

0.390244	0.666667	0.33758	0.342857	0.551724	0.615385	0.452991	0.5
0.372093	0.470588	0.33758	0.342857	0.457143	0.444444	0.438017	0.444444
0.355556	0.533333	0.333333	0.342857	0.761905	0.888889	0.569892	0.571429
0.347826	0.5	0.341935	0.342857	0.421053	0.571429	0.445378	0.5
0.380952	0.727273	0.33758	0.342857	0.666667	0.666667	0.504762	0.48
0.421053	1	0.350993	0.363636	0.842105	0.8	0.56383	0.5
0.347826	0.615385	0.33758	0.352941	0.888889	0.666667	0.697368	0.666667
0.380952	0.615385	0.341935	0.342857	0.592593	0.888889	0.576087	0.6
0.333333	0.444444	0.33758	0.342857	0.842105	0.8	0.616279	0.705882
0.372093	0.5	0.341935	0.333333	0.761905	0.888889	0.588889	0.571429
0.380952	0.470588	0.438017	0.48	0.5	0.888889	0.5	0.6
0.444444	0.470588	0.452991	0.461538	0.592593	0.571429	0.828125	0.75
0.390244	0.5	0.424	0.461538	0.551724	0.8	0.53	0.705882
0.592593	1	0.495327	0.5	0.695652	0.571429	0.588889	0.705882
0.432432	0.615385	0.445378	0.461538	0.571429	0.615385	0.552083	0.5

5 结束语

本文将传统的灰关系分析中参考(样本)向量改变为排序序号或者样本标号,将灰关系分析中的横向的时间域改变为垂直方向的特征域,计算每个特征对应列和参考向量之间的灰关系系数,获得每个特征和参考向量之间的灰色关联度,从而求解了各个特征向量和参考向量之间的相似性,从而可简单方便地进行特征的重要性排序.未来的工作将聚焦于如何应用计算所得灰色关联度指导后续的特征选择.

参考文献:

- [1] Tao Liu, Shengping Liu, Zheng Chen et al. A evaluation on feature selection for text clustering [C]// Proc of the 20 th International Conference on Machine Learning, Washington DC: Morgan Kaufmann Publishers, 2003: 488-495.
- [2] Yang Y, Pedersen J O. A comparative study on feature selection in text categorization [C]// Proc of 14 th International Conference on Machine Learning. Nashville: Morgan Kaufmann Publishers, 1997: 412-420.
- [3] Galavotti L, Sebastiani F, Simi M. Feature selection and negative evidence in automated text categorization [C]// Proceedings of the ACM KDD 2000 Workshop on Text Mining. Boston: ACM press, 2000: 287-290.
- [4] Dash M, Liu H. Feature selection for classification [J]. International Journal of Intelligent Data Analysis, 1997, 1(3): 131-156.
- [5] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.
- [6] Xiaoyan Shaoa, Xiang Li. Soil heavy metal pollution evaluation using grey relational analysis [M]. Energy Procedia: 2011(11):4878-4882.
- [7] Chih-Hung Tsai, Ching-Liang Chang, Lieh Chen. Applying grey relational analysis to the vendor evaluation model[J]. International Journal of the Computer, the Internet and Management, 2003, 11(3):45-53.
- [8] Ching-Liang Chang, Chih-Hung Tsai, Lieh Chen. Applying grey relational analysis to the decathlon evaluation model[J]. International Journal of The Computer, the Internet and Management, 2003, 11(3):54-62.
- [9] 生佳根, 刘思峰. 基于灰关系分析的模糊聚类[J]. 情报学报, 2010, 29(3):493-496.
- [10] Frank. A, Asuncion, A. (2010). UCI Machine Learning Repository[EB/OL][2012-05-06]. <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.

作者简介:



生佳根 男,(1962—),博士,教授.研究方向为知识工程和数据挖掘.



胡雨晴 女,(1989—),硕士研究生.研究方向为知识工程.