

DOI: 10.13382/j.jemi.2017.10.006

基于关联规则挖掘的气象观测设备一致性检测算法*

李涛 郁美辰 陆正邦 林陈 张灿

(南京信息工程大学 电子与信息工程学院 南京 210044)

摘要:为了保证气象观测设备采集数据的稳定性,从而需要对观测设备进行一致性检测。提出了一种基于兴趣度的关联规则的算法。并将该兴趣度关联规则挖掘算法应用于气象观测设备一致性检测上,可以形成关联规则气象观测设备一致性的模型。通过真实数据验证表明,该算法不仅能够挖掘出所有相关性很强的规则,同时与同类非 Apriori 类的算法相比,在时间性能上更加优越。通过该关联规则算法挖掘出所有关联项对形成范例库,利用规则匹配的方法对设备之间进行一致性检测,对算法实验优化,得到最优参数解,从而判定设备一致性。

关键词:气象观测设备;一致性检测;兴趣度;关联规则

中图分类号: TP181;TN81 **文献标识码:** A **国家标准学科分类代码:** 520.20

Consistency detection algorithm for meteorological observation equipment based on association rules mining

Li Tao Yu Meichen Lu Zhengbang Lin Chen Zhang Can

(School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: In order to ensure the stability of data acquisition by meteorological observation equipment, the consistency detection of the observation equipment is needed. An algorithm of association rules based on interest degree is proposed in this paper. The interest association rule mining algorithm is applied to the consistency detection of meteorological observation equipment, and the consistency model of the meteorological observing equipment can be formed. Through real data validation, it shows that the algorithm can not only excavate all the very strong rules, but also better in the time performance compared with the algorithm of non-Apriori class. Through this association rule algorithm, all the correlation items are excavated to form the example base, the rule matching method is used to detect the consistency between the equipment, the optimization of the algorithm is obtained, the optimum parameter solution is given, and the equipment consistency is determined.

Keywords: meteorological observation equipment; consistency detection; interest degree; association rules

0 引言

气象数据是整个气象观测环节的基础,气象数据的准确、实时是为气象部门做出决策的基本保障^[1,2]。为了保证气象观测数据的准确性、可比较性,在气象观测设备进入气象业务应用之前,必须根据功能规格要求书或其

他技术要求,进行设备一致性检测,来确保设备在观测业务系统的运行质量。因此,气象观测设备一致性检测是气象观测技术专用设备进入业务应用的重要环节。

传统的一致性检测主要利用概率统计中的正态总体均值的假设检验的方法^[3],对同一厂家的不同气象观测设备观测数据进行一致性检测。传统的一致性算法能较为准确地判断气象设备的一致性情况,但是传统的一致

收稿日期:2017-04 Received Date: 2017-04

* 基金项目:公益性行业(气象)科研专项(GYHY201306070)、江苏高校品牌专业建设工程(PPZY2015B134)、江苏省高等学校大学生创新创业训练计划项目(201610300031)资助

性检测数据必须满足正态分布以及检测阈值是基于历史资料得出的界限值^[4-5]。因此,传统的一致性检测的准确性以及时间效率不是很高。

对于两台同一厂家的同一类型的气象观测设备,放置在相同的环境下采集气象观测数据,气象设备采集数据中要素存在着一定的关联项^[6-7],提出了基于关联规则的气象观测设备一致性检测模型。而传统的关联规则挖掘算法都是基于 Apriori 算法优先挖掘出频繁项集^[8-9],从频繁项集中生成关联规则,传统的关联规则存在大量中间项集、难以发现低频信息等缺点^[10]。

针对传统关联规则算法中存在具有高频的频繁项集的不足,本文接着从相关性角度^[11-12]出发提出一种新的基于兴趣度的关联规则挖掘算法,通过项或项对超集的兴趣度上界来裁剪搜索空间。通过基于兴趣度的关联规则算法来挖掘气象数据并提取出设备数据所有关联项对形成范例库,来检测气象观测设备一致性^[13-14]。

1 基本理论

1.1 兴趣度定义

如果规则 $A \Rightarrow B$ 成立,则须满足条件:

$$p(B|A) > p(B)$$

物理意义表示在 A 出现的情况下 B 出现的概率要高于 B 直接出现的可能性,这样才能表现 A 促进 B 的出现。所以从相关性的角度出发,会考虑到一个新的兴趣度度量模型^[15]:

$$Interest(A \Rightarrow B) = \frac{p(B|A) - p(B)}{p(B|A) + p(B)} \quad (1)$$

该兴趣度度量有以下几个性质:1) 该度量具有上下界 $[-1, 1]$,能有效得控制参数的输入设置;2) 若兴趣度 $Interest = 0$,表示模式中的项集 A, B 之间相互独立;3) 若 $Interest > 0$,则表示模式中的项集 A 与 B 正相关;4) 如果兴趣度 $Interest < 0$ 即表示模式中的项集 A 与 B 负相关^[15]。

性质 1 已知模式 X , 兴趣度度量 $Interest(A \Rightarrow B)$ 具有上下界 $[-1, 1]$ 。

证明:

$$Interest(A \Rightarrow B) = \frac{p(B|A) - p(B)}{p(B|A) + p(B)} = \frac{p(AB) - p(A)p(B)}{p(AB) + p(A)p(B)} \leq 1 \quad (2)$$

$$Interest(A \Rightarrow B) = \frac{p(B|A) - p(B)}{p(B|A) + p(B)} = \frac{p(AB) - p(A)p(B)}{p(AB) + p(A)p(B)} \geq -1 \quad (3)$$

证毕。

性质 2 针对模式 $X = \{A, B\}$ 以及相关度度量

$Interest(A \Rightarrow B)$ 。

1) 若兴趣度 $Interest = 0$,则模式中的 A, B 相互独立;兴趣度 $Interest = 0$ 时,式(1)中的 A, B 相互独立时,则 $p(AB) = p(A)p(B)$,那么 $Interest(A \Rightarrow B) = 0$;

2) 若兴趣度 $Interest > 0$, $\frac{p(AB) - p(A)p(B)}{p(AB) + p(A)p(B)} > 0$, $p(AB) > p(A)p(B)$,则表示项集 A 与 B 正相关;

3) 若兴趣度 $Interest < 0$, $\frac{p(AB) - p(A)p(B)}{p(AB) + p(A)p(B)} < 0$, $p(AB) < p(A)p(B)$,则表示项集 A 与 B 负相关。

1.2 超集兴趣度上界

性质 3 有一项集 X ,则相关性度量如下:

$$M(sup(x), sup(y), sup(z)) = M(sup(x), sup(y), sup(x \setminus y)) = corr_{A,B} = \frac{sup(AB)}{sup(A)sup(B)} = sup(x) / [sup(y) \cdot sup(x \setminus y)] \quad (4)$$

式中: $z = x \setminus y$ 。

关于相关度度量的定理如下。

定理 1 当 $sup(y), sup(z)$ 以及事务数据数 n 保持不变时,相关度度量 $M(sup(x), sup(y), sup(z))$ 与 $sup(x)$ 成正比关系。

证明:

$$M(sup(x), sup(y), sup(z)) = sup(x) / [sup(y) \cdot sup(z)] = sup(x) / [sup(y) \cdot sup(x \setminus y)]$$

因此,当 $sup(y), sup(z)$ 以及 n 保持不变时,相关度度量 $M(sup(x), sup(y), sup(z))$ 正比于 $sup(x)$ 。

定理 2 若 $sup(x)$ 以及 $sup(z)$ (或者 $sup(y)$) 不变时,相关度度量 $M(sup(x), sup(y), sup(z))$ 与 $sup(y)$ (或者 $sup(z)$) 反比关系。

证明:

当其他参数 $sup(x), sup(z)$ 保持不变时, $M(sup(x), sup(y), sup(z))$ 只与参数 $sup(y)$ 有关,即两种是反比关系。

针对上述相关度度量定理,可获得关于相关度度量的上限值;给定一个项集 X ,以及项集超集 $x, x \subseteq X$,则项集 X 的超集 x 的相关度度量的上限值为^[15]:

$$M(sup(x), sup(x), \max_{i \in x}(sup(\{i\}))) = \frac{1}{\max_{i \in x}(sup(\{i\}))} \quad (5)$$

依据如下:

1) $sup(x) \leq sup(x)$,相对于其他参数,若度量中的第一个参数 $sup(x)$ 最大时,则相关度度量 M 最大。

2) 当 $sup(y)$ 与 $sup(z)$ 最小时 M 最大,而 $sup(y), sup(z)$ 都不小于 $sup(x)$,而此时 $sup(x)$ 设置为 $sup(x)$,则当 $sup(y) = sup(x)$ 时, M 最大。

3) 当 $sup(y) = sup(x)$, 则 $z = x \setminus y = x \setminus x$, 其中 $\{j\} = x \setminus x$, 而项是以支持度从小到大排序, 则 $sup(\{j\}) \geq \max(sup(\{i\}))$, 其中 $i \in x$ 。所以当相关度量 M 的第 3 个参数最小时, 即为 $\max(sup(\{i\}))$, M 则最大, 因此该最大值就是项集相关度量的上限值^[15]。

性质 4 所以项对 x 的超集的相关度量的上限值为 $\frac{1}{\max(sup(\{i\}))}$, 进一步可求每一个项的超集的相关度上界为 $\frac{1}{sup(\{i\})}$ ^[15]。

1.3 兴趣度与相关度量关系

兴趣度量:

$$Interest(A \Rightarrow B) = \frac{p(B|A) - p(B)}{p(B|A) + p(B)} = \frac{\frac{p(AB)}{p(A)p(B)} - 1}{\frac{p(AB)}{p(A)p(B)} + 1} = \frac{corr_{A,B} - 1}{corr_{A,B} + 1} = 1 - \frac{2}{corr_{A,B} + 1} \quad (6)$$

由式(6)发现, 兴趣度与相关度成正比关系。则假设有一项集 X , 分割成 y 和 z 两部分, 则项集 X 的超集 x 的兴趣度量的上界在相关度最大时取值, 即:

$$Interest_{max} = 1 - \frac{2}{M + 1} = 1 - \frac{2}{\max(sup(\{i\})) + 1} \quad (7)$$

1.4 项对超集的兴趣度上界应用

算法将对兴趣度上界的搜索空间进行裁剪, 即当项 $\{i\}$ 或项对 x 的超集的兴趣度上界小于兴趣度阈值 t 时, 则可裁剪项 $\{i\}$ 或项对 X 搜索空间, 降低了算法的复杂度^[15]。

2 关联规则气象观测设备一致性模型

关联规则气象设备一致性模型在孤立点分析和异常检测中的应用出发, 结合气象数据的特点以及误差等作为背景。

基于关联规则的气象观测设备一致性检测模型主要包含了 3 大步骤, 即数据预处理、产生关联规则、规则匹配。具体流程如图 1 所示。

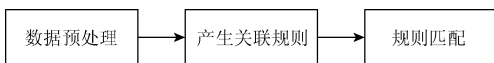


图 1 气象观测设备一致性检测模型

Fig. 1 The consistency detection model for meteorological observation equipment

基于关联规则的气象观测设备一致性检测模型分为 3 大步, 每一步中又有各自的主要处理过程。模型的输

入, 两台同一厂家相同型号的气象观测设备采集数据。输出, 两台设备一致性检测结果。

1) 将两台设备分别作为标准设备和被测设备。将标准设备的各要素数据进行离散化处理, 得到离散型的数据。

2) 将基于兴趣度的关联规则算法对预处理后的数据进行挖掘, 得到频繁项集, 通过项集产生关联规则库。

3) 将待测设备数据也进行预处理, 然后将其的每条记录和关联规则库进行匹配, 得到异常记录值。

3 实验结果和分析

3.1 数据选取

气象观测设备是准确提供气象数据的基础, 气象观测设备是保证气象数据的高质量, 保证气象数据准确性, 具有重大的实际意义。因此保证气象观测设备一致性非常重要, 现如今, 关联规则挖掘算法在气象领域中的应用越来越多。

本文采用 2015 ~ 2016 年某测试场地某种测试设备采集的数据进行检测气象观测设备一致性检测。气象设备采集的气象数据要素包括气温、相对湿度、降雨量、两分钟平均风速、极大风速等作为考察对象, 这些要素都是以每小时间隔进行采集, 具体数据如表 1 所示。

表 1 原始数据集

Table 1 Original data sets

气温/ ℃	相对湿度/%	降水量/ mm	2 min 平均风速/ (m · s ⁻¹)	...
-0.5	35	0	0.8	...
-1.1	36	0	1.4	...
...
5.4	91	1.1	3	...
...

3.2 数据处理

由于关联规则目前无法处理量化的连续数据, 因此, 需要对每个气象要素进行数据变换, 将连续数据进行离散化, 按照每个要素进行数据变换, 将连续数据进行离散化, 按照每个要素的属性值划分为不同的区间。区间划分的标准划分为 8 段。下面按照区间段 8 段划分, 数据预处理结果如表 2 所示。

3.3 匹配过程

气象观测设备采集的气象数据经过数据预处理后, 应用基于兴趣度的关联规则算法应用到气象观测设备一致性检测中。

表 2 数据集格式变换

Table 2 Data set format transformation

气温/ ℃	相对湿度/%	降水量/ mm	2 min 平均风速/ ($m \cdot s^{-1}$)	...
q2	x2	j1	c2	...
q2	x2	j1	c2	...
...
q3	x5	j2	c4	...
...

1) 将一个厂家给出的相同的两台设备进行 A、B 编号。先对厂家 A、B 两台设备得到的数据预处理,然后将设备 A 经过预处理的数据用新的关联规则的算法进行挖掘,得到所有关联规则对组成规则库,并作为一组范例库。

2) 将设备 B 观测数据作为待测数据的每条记录和规则集中每一条规则进行匹配,用以判断记录是否异常,直到出现 u 条不匹配的规则,则定位该记录为异常。

3) 将设备 B 观测数据经过关联规则挖掘作为另外一组范例库,将设备 A 观测数据作为待测数据与设备 B 挖掘出来的范例库进行匹配,找到设备 A 中的异常记录。

4) 将设备 B 采集数据作为待测数据得到的异常数据与设备 A 采集数据作为待测数据得到的异常数据进行比较,如果两者异常数据条数的差值在规定阈值以内,那么就说明气象观测设备一致。

3.4 实验过程

气象观测设备一致性检测对同一场地同一厂家同一型号的两台设备进行测试,设备 A、B 分别作为待测数据进行实验,通过对检测率和误报率这 2 个参数来衡量。

检测率是指检测出的异常记录的的概率,检测率 = 检测出的异常记录个数/测试数据中所有的异常记录个数。

误检率是检测到的记录中不是异常记录的个数占总的异常记录的百分比,误检率 = 检测到的记录中不是异常记录的个数/检测到的所有记录个数。

理想状态下,检测数达到 100%, 而误检数为 0。

首先从厂家中的标准设备的正常数据,随机选择部分数据组成多组测试数据并且手动植入误差数据。然后将带有误差数据的测试数据与标准规则集中的规则进行匹配,通过几组仿真实验来优化控制参数达到最优检测效果,如表 3~6 所示。

表 3 第一组测试数据的数据结构

Table 3 Data structure of the first set of test data

q4	l3	s2	x7	...
q4	l3	s2	x5	...
...
q4	l6	s3	x8	...
q5	l6	s3	x6	...
...

表 4 第 1 组测试数据不同参数下的结果对比

Table 4 Comparison of the results of the first set of test data under different parameters

t	u	所有记录	包含的异常记录	检测率	误检率
0.4	≥ 16	25 个	18 个,包含 1~18 行异常数据	1	0.28
	≥ 18	23 个	18 个,包含 1~18 行异常数据	1	0.22
	≥ 20	18 个	18 个,包含 1~18 行异常数据	1	0
	≥ 21	17 个	17 个,缺少第 13 行异常数据	0.94	0
	≥ 22	12 个	12 个,只有 1~12 行异常数据	0.67	0
	≥ 23	12 个	12 个,只有 1~12 行异常数据	0.67	0
0.5	≥ 12	20 个	18 个,包含 1~18 行异常数据	1	0.1
	≥ 14	18 个	18 个,包含 1~18 行异常数据	1	0
	≥ 16	9 个	9 个,包含第 4~10、15、16 行数据	0.5	0
0.6	≥ 5	21 个	18 个,包含 1~18 行异常数据	1	0.14
	≥ 6	18 个	18 个,包含 1~18 行异常数据	1	0
	≥ 7	12 个	12 个,包含第 4~10、14~18 行异常数据	0.67	0

表 5 第 2 组测试数据的数据结构

Table 5 The second set of test data structures

q5	l4	s2	x4	...
q5	l4	s2	x3	...
...
q3	l5	s2	x7	...
...

表 3、5 分别是第 1 组和第 2 组数据结构说明,表 4、6 是分别对数据在不同参数下的结果对比。通过 2 组仿真实验发现,当参数匹配度 $u = 14$, 兴趣度阈值 $t = 0.5$ 时,检测效果最好,基本能将所有异常记录都能检测出来。

当设置好参数后,就可以对厂家生产的设备进行测试其一致性。

3.5 实验结果分析

经过上面多组实验对参数进行的优化,得到最优参数 $t = 0.5, u = 14$ 。

设备 A、B 采集的数据都是同一时间点的的数据,数据格式以及数目相同。这样,将设备 A 挖掘出的规则集作为范例库,B 设备的数据一一和规则集中规则进行匹配,发现 B 设备中的数据有 82 条异常记录,而同样将设备 B

表6 第2组数据不同参数下的结果对比
Table 6 Comparison of the results of the second set of data under different parameters

t	u	所有记录	包含的异常记录	检测率	误检率
0.4	≥ 16	19个	12个,包含1~12行异常记录	1	0.37
	≥ 18	17个	12个,包含1~12行异常记录	1	0.29
	≥ 20	6个	6个包含第2~6以及8行数据	0.5	0
	≥ 12	14个	12个,包含1~12行异常记录	1	0.14
0.5	≥ 14	12个	12个,包含1~12行异常记录	1	0
	≥ 16	11个	11个,缺少了第7行数据	0.92	0
	≥ 18	4个	4个,只有第3~6行异常数据	0.33	0
	≥ 5	14个	12个,包含1~12行异常记录	1	0.14
0.6	≥ 6	12个	12个,包含1~12行异常记录	1	0
	≥ 7	5个	5个,只有2~6行异常数据	0.42	0
	≥ 8	4个	4个,只有3~6行异常数据	0.33	0

挖掘出的规则集作为范例库,A设备的数据一一和规则集中规则进行匹配,发现A设备中的数据有108条异常记录。两组设备异常数据数目的差值为26条,差值范围比较小,两台设备的一致性较好。

对厂家2中的设备C以及设备D,同样以上述的步骤来验证厂家2的设备一致性。设置参数 $t=0.5$ $u=14$,将设备C挖掘出的规则集作为范例库,D设备的数据一一和规则集中规则进行匹配,发现D设备中的数据有281条异常记录,而同样将设备D挖掘出的规则集作为范例库,C设备的数据一一和规则集中规则进行匹配,发现C设备中的数据有506条异常记录。两台设备异常数据的差值为225条,差值范围较大,即设备C与设备D的一致性较差。

4 结论

本文针对气象观测设备一致性检测的需求,从关联规则挖掘算法在孤立点的分析应用出发。通过与传统的质控方法对比发现,提出了新的基于兴趣度的关联规则算法,利用它对气象观测设备的一致性进行检测,该检测方法在精度和灵敏度方面都得到了很大提高,通过实验仿真验证了方法的可行性。依据这套检测算法,准确而又快速地找到异常气象数据,对气象观测设备的一致性

检测具有很高的实际应用价值。

参考文献

- [1] 彭宇, 庞景月, 刘大同, 等. 大数据:内涵、技术体系与展望[J]. 电子测量与仪器学报, 2015, 29(4): 469-482.
PENG Y, PANG J Y, LIU D T, et al. Large data: connotation, technical system and prospect [J]. Journal of Electronics Measurement and Instrumentation, 2015, 29(4): 469-482.
- [2] 刘艳中, 敖振浪. 一种新型自动气象站外接数据存储系统的设计[J]. 电子测量技术, 2015, 38(8): 109-112.
LIU Y ZH, AO ZH L. Design of a new automatic weather station external data storage system [J]. Journal of electronic measurement technology, 2015, 38(8): 109-112.
- [3] 李荣华, 徐九韵. 多维正态总体零均值的假设检验[J]. 中国石油大学学报:自然科学版, 2001, 25(6): 112-113.
LI R H, XU J Y. Hypothesis test of multi-dimensional normal population zero mean [J]. Journal of China university of petroleum science, 2001, 25(6): 112-113.
- [4] 王海军, 刘莹. 综合一致性质量控制方法及其在气温中的应用[J]. 应用气象学报, 2012, 23(1): 69-76.
WANG H J, LIU Y. Integrated consistency quality control method and its application in air temperature [J]. Journal of Applied Meteorology. 2012, 23(1): 69-76.
- [5] 张楚金, 王耀南, 卢笑, 等. 基于假设验证和改进HOG特征的前车检测算法[J]. 电子测量与仪器学报, 2015, 29(2): 165-171.
ZHANG CH J, WANG Y N, LU X, et al. Forward vehicle detection algorithm based on hypothesis verification and improved hog characteristics [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(2): 165-171.
- [6] 闫少华. 离群点挖掘在入侵检测中的研究与应用[D]. 广州:广东工业大学, 2011.
YAN SH H. Outliers mining research and application in intrusion detection [D]. Guangzhou: Guangdong university of technology, 2011.
- [7] 郑忠平. 基于关联规则和聚类分析的异常天气挖掘[D]. 成都:电子科技大学, 2011.
ZHENG ZH P. Based on the mining association rules and clustering analysis of abnormal weather [D]. Chengdu: University of electronic science and technology, 2011.
- [8] 陈宁军, 高志年. 一种改进的正负关联规则挖掘算法[J]. 计算机科学, 2011, 38(12): 191-193.

- CHEN N J, GAO ZH N. An improved mining algorithm of positive and negative association rules [J]. Computer Science, 2011, 38(12):191-193.
- [9] 钱光超, 贾瑞玉, 张然, 等. Apriori 算法的一种优化方法[J]. 计算机工程, 2008, 34(23):196-198.
- QIAN G CH, JIA R Y, ZHANG R, et al. Apriori algorithm is an optimization method [J]. Computer Engineering, 2008, 34(23):196-198.
- [10] 吕杰林, 陈是维. 基于相关性度量的关联规则挖掘[J]. 浙江大学学报:理学版, 2012, 39(3):284-288.
- LU J L, CHEN SH W. Based on the correlation measure of association rule mining [J]. Journal of Zhejiang University: Science Edition, 2012, 39(3):284-288.
- [11] XIONG H, SHEKHAR S, TAN P N, et al. Exploiting a support-based upper bound of Pearson's correlation coefficient for efficiently identifying strongly correlated pairs [C]. Proceedings of the tenth ACM SIGKDD international conference on Knowledge Discovery and Data Mining, ACM, 2004:334-343.
- [12] 钱铁云, 冯小年, 王元珍. 超越支持度-置信度框架的负相关对规则挖掘[J]. 计算机科学, 2005, 32(10):124-127.
- QIAN T Y, FENG X N, WANG Y ZH. Beyond support-confidence framework is negatively related to rule mining [J]. Journal of Computer Science, 2005, 32(10):124-127.
- [13] 蒋忠凌, 廖俊必, 黄玉波, 等. 基于机器视觉的化成箔缺陷在线检测系统设计[J]. 电子测量技术, 2015, 34(7):27-32.
- JIANG ZH L, LIAO J B, HUANG Y B, et al. Online to foil defect detection based on machine vision system design [J]. Journal of Electronic Measurement Technology, 2015, 34(7):27-32.
- [14] 颜佩, 丁亚军, 钱盛友, 等. 基于 K 均值聚类的组织损伤等级判定研究[J]. 电子测量与仪器学报, 2017, 31(3):468-473.
- YAN P, DING Y J, QIAN SH Y, et al. Study on grading of tissue damage based on K means clustering [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(3):468-473.
- [15] 李涛, 林陈, 王丽娜. 一种改进的相关项对挖掘算法研究[J]. 计算机仿真, 2016, 33(8):223-228.
- LI T, LIN CH, WANG L N. An improved algorithm for mining related items to study [J]. Computer Simulation, 2016, 32(8):223-228.

作者简介



李涛, 分别在 1999 和 2002 年于南京信息工程大学获得学士学位和硕士学位, 在 2009 年于南京航空航天大学获得博士学位, 现为南京信息工程大学副教授, 主要研究方向为数据挖掘、信息处理等。

E-mail: lthnxx@21cn.com

Li Tao received B. Sc. and M. Sc. from Nanjing University of Information Science and Technology in 1999 and 2002, and Ph. D. from Nanjing University of Aeronautics and Astronautics in 2009. Now he is associate professor in Nanjing University of Information Science and Technology. His main research interest includes data mining, and information processing.



郁美辰 (通讯作者), 2015 年于南京信息工程大学获得学士学位, 现为南京信息工程大学硕士研究生, 主要研究方向为数据挖掘、信息处理。

E-mail: ymc_nuist@163.com

Yu Meichen (Corresponding author) received B. Sc. from Nanjing University of Information Science and Technology in 2005. Now she is a M. Sc. candidate in Nanjing University of Information Science and Technology. Her main research interest includes data mining, and information processing.