基于SMOTE算法与决策树的沙尘暴短期预警研究

张振华[a],徐瑾辉[b],李龙欣[b],马 超[b],黄江楠[a],谢政宏[c]

(广东外语外贸大学 a.经济贸易学院;b.金融学院;c.思科信息学院,广东 广州 510006)

摘要:针对沙尘暴灾害发生时间预测准确率较低、传统的预测模型预测效果欠佳问题,建立了基于SMOTE算法与决策树算法的沙尘暴预测模型.该模型利用西北六省的气象观测数据,较好地解决了稀有类的分类问题,总体预测成功率达到76.25%.研究结果表明该模型分类准确率高、泛化性能好、抗噪音、鲁棒性好,较好地解决了沙尘暴预测中不平衡样本的分类预测问题,可用于实际的沙尘暴预警.

关键词:沙尘暴预测;稀有类分类;SMOTE;决策树

中图分类号:R284.2 **文献标志码**:A **文章编号**:1674-358X(2015)03-0040-07

近几十年来,沙尘暴灾害在我国的甘肃、新疆、宁夏等西北地区频繁发生,影响甚至危害到当地的经济发展、生活与环境[1].至今关于沙尘暴灾害发生的时空规律的研究表明沙尘暴灾害的发生具有一定的规律性,其发生频率大体由西北向东南方向逐步减少.沙尘暴灾害的发生与当地气候的变化和地面沙尘物质的消长有着密切的关系:在当地气候温暖、潮湿的时候,沙尘暴灾害发生的频率比较低;相对而言,在寒冷、干燥的气候时期,沙尘暴灾害发生的频率比较高[2].沙尘暴灾害通常在雨季到来之前出现,而根据已有的研究表明,世界上绝大部分地区的沙尘暴灾害主要发生在春季,在我国的次数将近全年的一半,而强沙尘暴天气也主要在这个季节发生[3].我国沙尘暴灾害在4月发生的频率最大[4],但由于复杂的气候原因,近年其发生有提前的趋势,在9~10月发生的频率为全年最低[5-6].

关于气候因素对沙尘暴灾害频率格局的影响,国内外学者以降水、风速等气候因子建立了综合气候影响模型,并分析出其对沙尘暴灾害发生频率的影响[7-9].在我国,张小玲等[10]指出,北京地区近几年沙尘暴灾害发生频次的变化主要受气温变化、降水量等原因影响;李岩瑛等[11]指出沙尘暴灾害的长期预报取决于冬春两季的气温、降水量和大风日数等气候因子数据,中期预报需参考国内外相关预报数据,短期预报与大气环流条件、分型指标有关,而短时临近预报,也称超短期预报,与当地的高空大风形势、地面上游有无大风沙尘暴天气有着密切的关系;黄富祥等[12]建立的毛乌素沙地气候特征的定量模型利用气候影响指数拟合了沙尘暴频率;张杰等[13]采用沙尘暴灾害发生时间、冬季降水量等数据,研究了我国西北地区东部冬季的降水与次年沙尘暴灾害发生的关系;李锡福[14]、刘立超等[15]分别分析了青海、宁夏地区的天气气候特征及沙尘暴灾害成因;赵建华等[16]指出气温、冻土冻深与沙尘暴的负相关关系;赵光平等[17]研究了沙尘暴灾害发生发展规律与生态退化的关系;郭铌等[18]研究了我国西北地区地形、地势与沙尘暴灾害发生的关系;矫梅燕等[19]分析了2002年、2003年典型沙尘天气过程和冷空气过程中大气动力条件的机理作用;许炯心[20]研究发现我国黄土高原地区自然地理因子与沙尘暴灾害的发生密切相关;至于在沙尘暴的预测工作方面,李登科等[21]认为陕西是否出现沙尘暴天气,其中重要一点就是临近陕西的上游地区有无沙尘暴出现;钱正安等[22]指出华北平原、内蒙古草原东南部、鄂尔多斯高原、阿拉善高原、塔里木盆地为我国五大沙尘暴中心;胡金明等[23]、Dong等[24]的研究表明我国沙尘天气的多发地带是北方农牧交错带、沙漠边缘带、沙漠—绿洲过渡带,西北地区强沙尘暴高发区则为吐鲁番、和田和民勤;刘景涛等[25]指出内蒙古的朱日和地区因常受西北路冷空气和强西路冷空气的影响,成为华北强与特强沙尘暴的中心地区.

收稿日期:2015-05-17

基金项目:广东省自然科学基金项目(2014A030313575);2015年广东大学生科技创新培育专项资金一般项目(308-GK151013)

作者简介:张振华(1972—),男,湖南郴州人,副教授,博士,主要从事模糊识别与模糊推理研究.

在这里需要指出的是,尽管相关领域的学者对沙尘暴产生前后的物理机制有着非常广泛的研究,但均存在着一个严重的不足,即很难将已有的研究成果运用到实际的沙尘暴预测中,无法准确地预测出沙尘暴灾害什么时候会发生.故而,本文根据前人成果,综合了一些与沙尘暴灾害发生密切相关的指标,对沙尘暴灾害的发生条件进行研究,试图将沙尘暴预测成为现实,以期对沙尘暴的长期发展趋势做出良好的判断,同时对沙尘暴灾害进行高精度的预测与预警.

1 理论模型与假设

1.1 SMOTE 算法

对于稀有类问题来说最简单直观的做法就是改变各类的分布情况,将存在的不平衡的训练样本预处理为一般的训练样本,将稀有类分类问题转化为普通分类问题.SMOTE(synthetic minority over-sampling technique)算法是应用在不平衡数据集学习的一种新方法,是一种改进的过度采样策略,具体就是指通过对稀有类样本的人工合成来提高稀有类样本的比例,同时这种方法也降低了样本数据的过度偏斜问题.SMOTE 算法的特点[26]在于不是简单地按照随机过采样方法进行复制样例,而是增加一些新的并不存在的样例来达到样本平衡,因此它可以在一定程度上避免分类器过度拟合.

假设有少数类样本,每一个样本 x,搜索其 k(通常取 5)个少数类最近邻样本;若向上采样的倍率 N,则在其 k 个最近邻样本中随机选择 N 个样本,记为 $y_1,y_2,\cdots,y_n$;在少数类样本 x 与 y_j($j=1,2,\cdots,n$)之间进行随机线性插值,构造新的少数类样本 p_j,即

$$p_j = x + \mathrm{rand}(0,1) \times (y_i - x), j = 1,2,\cdots,n,$$

式中 rand(0,1)表示区间(0,1)内的一个随机数.将这些新合成的稀有类样本点加入并合并到原来的数据集里即产生了新的训练集.

图 1 描述人工合成新数据这一过程,x_i 为某一个少数类实例,x_{i1}、x_{i2}、x_{i3}、x_{i4} 分别为 x_i 的 4 个近邻,r_1、r_2、r_3、r_4 为 4 个新生成的人造数据.通过 SMOTE 算法生成新的稀有类数据,增加了通用性,不会发生如精确复制样本一样引起的过度拟合的问题.

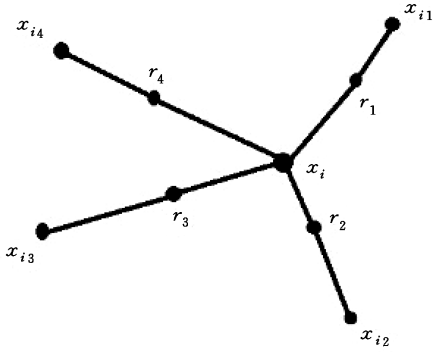

图 1 SMOTE 算法人工合成新数据过程

1.2 决策树算法

决策树(decision tree)算法是一种实现分治策略的层次数据模型,是一种判断可行性的决策分析方法.在数据挖掘、统计学习、机器学习及模态识别中,决策树是一种用于监督学习的层次模型,代表了对象属性与对象值之间的一种映射关系.

根据所定的样本集 L,由以下 3 个步骤构建分类树:

1)低规划分训练样本:使用 L 构建树 T_{max},使得 T_{max} 中每一个叶节点要么非常小(即给定值 N_{min} 远远大于节点内部所包含的样本数量);得到的惟一属性向量作为分支选择,或者纯节点(节点内部样本 Y 仅仅包含一类).

2)使用修剪算法构建一个有限的递减(节点数目)有序子树序列.

3)使用评估算法从第 2)步产生的所有子树序列中选出一棵最优树,该树即为最终的决策树.

由于 SMOTE 对于决策树算法的优化效果最明显,因此利用它对沙尘暴进行预警.本文使用的是由 JA-

VA 实现的 C4 .5 决策树—J48 .其优点是产生的分类规则易于理解,准确率较高 .

1 .3 基本假设

假设一:沙尘暴灾害的发生是可以提前预测的 .

假设二:所有的观测数据都是准确、真实的 .

假设三:所有地区沙尘暴灾害的形成机理都相同 .

假设四:沙尘暴灾害的形成不是人为主观故意的,只与自然因素有关 .

2 数据收集与指标选取

2 .1 数据来源

基于我国西北六省沙尘暴观测点的观测数据,来源于《中国强沙尘暴序列及其支撑数据集》(中国气象局国家气象中心资料室,2008 年).该数据集统计处理了 1954—2007 年我国各观测站沙尘暴出现时间、结束时间以及相关的大风、能见度等资料[27] .由于条件的种种限制,这里只选取 1961—2005 年西北六省的 193 个观测点的数据 .

2 .2 数据描述

使用的数据包含如下内容:

1)西北六省的 193 个沙尘暴观测点的编号及其地点名称 .

2)其中绝大部分观测点给出了 1961—2005 年间各个月份与沙尘暴相关的气象数据的观测值,即月平均风速、月大风发生、月平均气温、月降水量、月平均相对湿度、月沙尘暴天数 .

2 .3 指标选取

2000 年,中国科学院地质部门发布了题为《关于我国华北沙尘天气的成因与治理对策》[28] 的报告 .该报告根据过去几十年的观测数据,指出未来的几十年里,中国范围内降水量将会发生变化,气温显著上升,地表的蒸发量增加 .这些变化会导致土壤湿度的下降,形成沙尘暴发生的必要气候条件 .报告得出的结论是中国北方将会出现明显的干旱趋势 .张仁健等[29] 发现沙尘暴发生的物理机制十分复杂,指出该机制涉及沙源、强风以及大气层这 3 个条件,具体来说就是风、大风发生日数、降水量、蒸发量、湿度以及气温等 .李宁等[30] 探讨了多致灾因子对 Copula 联合分布模型在三维多致灾因子综合分析中的扩展,针对大风气候、丰富的沙尘源分布和不稳定的大气层这 3 个形成沙尘暴灾害的基本条件,以内蒙古镶黄旗 1990—2008 年的强沙尘暴灾害事件为案例,建立了经向环流指数、地面平均最大风速和地表土壤湿度 3 个基本特征变量的联合分布,计算了基于联合分布的联合重现期,其中三维 Frank Copula 在中高尾部分的拟合有很大提高 .

根据以上的研究以及 1961—2005 年中国强沙尘暴序列及其支撑数据集,将以下指数选为本文沙尘暴预测的必要指数,即月平均风速、大风日数、月平均温度、月降水量、月蒸发量、月平均相对湿度、沙尘暴发生日数 .

3 沙尘暴短期预警模型

3 .1 问题的背景与难点

沙尘暴的短期预警难点在于必须综合考虑各种影响沙尘暴发生的因素 .沙尘暴灾害是一种成因非常复杂的气象灾害,它的形成过程以及严重程度不仅与当地的地理环境有关(如地形地势、植被覆盖、沙源状况),亦与一些气象因素有关,如风、平均气温、降水、空气湿度等因素 .由于短期中一个地区的地理环境不会发生明显改变,因此假定只考虑气象因素(当然,考虑到地理环境的差异,分地区进行预测是必要的).在短期预测中,由于年份自变量保持不变,必须要依靠其他的气象数据进行预测 .

3 .2 数据描述与归纳

3 .2 .1 数据描述

考虑到 1981—2005 年数据较先进和完整,故而在采用它们的同时,考虑不同地区之间地理环境的差异,并以甘肃省内的站点为例(共 6900 个样本),建立短期预测模型 .

3 .2 .2 数据归纳

沙尘暴灾害的产生主要是受三大气候因子支配,它们分别是热力动力不稳定因子、强风因子和沙尘源因子 .所使用的数据中,包含了月平均风速、月大风发生、月平均气温、月降水量、月平均相对湿度、月蒸发量、月

沙尘暴天数的观测值.其中,热力动力不稳定因子与该地区月均温、月相对湿度有关;强风因子与该地区月均大风日数、月均风速有关;沙尘因子与该地区月降水量、月蒸发量、月相对湿度有关.综上所述,本文的数据基本覆盖了 3 个支配沙尘暴产生的重要因素,可以用来建立预测模型.

通过观测甘肃省沙尘暴的月发生天数,在共计 6900 个样本中,0 样本(即不发生沙尘暴的样本)的比例达到了 80% 以上,而发生沙尘暴的样本不足 20% .从概率论的角度而言,沙尘暴的发生是一个小概率事件;从统计样本的角度而言,沙尘暴的发生是一个稀有类[31].这种稀有现象使得传统的所有预测模型都失效了,因而提出新的统计预测模型.

3.3 基于 SMOTE 算法的样本平衡

根据已有的研究成果,处理稀有类的分类问题有以下的 2 条思路[32]:1)从训练集出发,通过一定的采样方法,改变训练集样本分布,降低不平衡程度,将稀有类分类问题转化为普通分类问题,再使用普通的统计学习方法来进行分类;2)从统计学习算法出发,改进算法使之适应不平衡分类问题,例如集成学习法和 EPRC 分类算法.此外,基于代价敏感学习的分类方法也是当前的研究热点之一.本文将结合以上 2 种思路,一方面平衡样本分布,另一方面使用集成学习以提升分类器性能.

3.4 短期预测实例

将前文提到的 SMOTE 算法与决策树算法结合,并以甘肃省为例,进行月沙尘暴天数的短期预测.

STEP 1:以甘肃省 1981—2005 年的沙尘暴相关气象数据为研究对象,有 6899 个样本,共包含了甘肃省各观测站点 25 年来的月度观测数据.包含的项目是年份、月份、月均风速、大风日数、月均气温、月降水量、月蒸发量、月均湿度、沙尘暴天数.其中,前 8 个项目为样本特征,沙尘暴天数为样本分类标签,是取值范围为 0～9的整数,记为第 0～9 类.在此,将其作为分类变量进行识别.以其中 4 个样本为例,样本格式见表 1.

表 1 样本格式

区站号	年	月	月均风速	大风日数	月均气温	月降水量	月蒸发量	月均湿度	沙尘暴天数
52323	1981	1	4.2	3.3	−12.0	1.4	50.2	52	0
52323	1981	3	4.5	4.6	−0.7	0.9	179.9	33	0
52323	1981	4	4.2	6.7	6.6	2.3	287.8	31	0
52323	1981	5	5.2	4.3	12.0	1.7	484.0	18	0

STEP 2:对这 6899 个样本随机分层抽取出 500 个样本作为训练集,另随机分层抽取 500 个样本作为测试集.样本分布 V 如图 2 所示.在图 2 中,从左起分别为分层抽取的 500 个样本中属于第 0,1,2,…,9 类标签的样本数目.

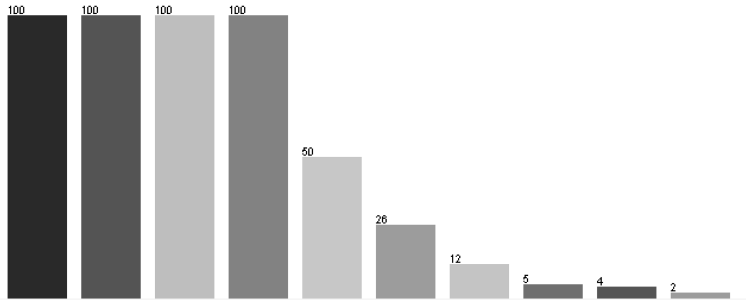

图 2 样本分布示意图

STEP 3:利用 SMOTE 算法对训练集进行处理,对于稀有类(第 5,6,7,8,9 类)进行人工合成操作,在保证一定合成样本质量的前提下,平衡样本分布.平衡后新训练集的分布如图 3 所示.

STEP 4:对经 SMOTE 算法处理后的新训练集,将样本的 8 个特征属性作为 J48 算法的输入,样本的月

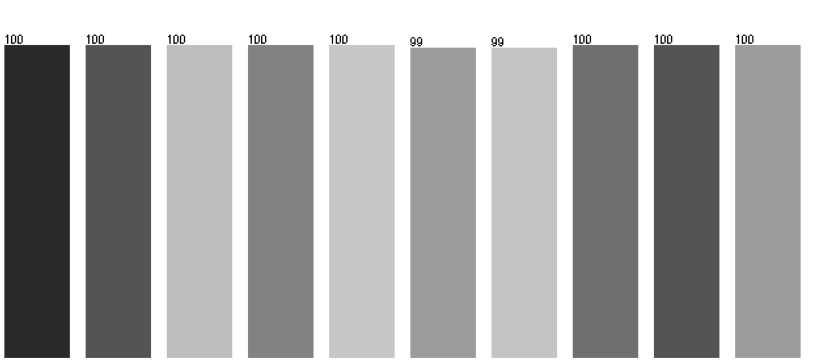

图 3　平衡后新训练集的分布示意图

沙尘暴发生天数作为输出,进行集成学习.在这里,并不对样本数据进行归一化处理,因为发现归一化反而会降低 J48 算法的性能.同时,得益于 J48 算法的特性,也不需要对样本的特征属性进行预处理或筛选.在这种情形下,将会发现 J48 算法表现依然非常优异.

STEP 5:对上一步训练得到的 J48 分类器,用包含 500 个样本的测试集对其泛化能力与预测能力进行测试.

STEP 6:实际应用.对于特定的沙尘暴观测站点,通过测出当前的各项气候因素值(如月份,风速,大风发生次数,降水量,温度等),输入 J48 模型得到相应的月沙尘暴天数的预测.以此为基础,可以定义沙尘暴的风险度,即沙尘暴天数预测值越高,则当前沙尘暴发生的风险越大,应提早预警并防范.

使用 SMOTE 与 J48 组合算法对沙尘暴进行短期预测,预测输出结果如图 4 所示.

```
=== Summary ===

Correctly Classified Instances         350               76.2527 %
Incorrectly Classified Instances       109               23.7473 %
Kappa statistic                          0.7146
Mean absolute error                      0.0631
Root mean squared error                  0.1888
Relative absolute error                 35.0599 %
Root relative squared error             62.9287 %
Total Number of Instances              459
```

图 4　决策树算法输出结果

结果表明,使用 SMOTE+J48 组合算法对沙尘暴进行短期预警,其准确率高达 76.25%.这说明该模型具有良好的预测效果.

3.5　预测效果对比

为了说明 SMOTE+J48 算法具有较其他算法更好的性能,选择了一些算法并与之进行对比.由于篇幅有限,只列出实验结果,具体见表 2.

表 2　本算法的沙尘暴短期预测模型输出结果与其他算法的对比

算法名称	SMOTE+J48	J48	SMOTE+Logistic	SMOTE+Decision Stump	SMOTE+Decision Table	SVM	BP ANN	Fuzzy ANN
准确率/%	76.25	48.37	26.58	17.43	32.46	32.46	51.85	54.28

经过对比可以看出,较之传统预测方法,本预测模型预测准确率非常高,可以用于实际的沙尘暴预警.

3.6　模型优点

所提出的基于 SMOTE 与决策树算法的沙尘暴预测模型有以下优点:1)较好地解决了沙尘暴预测这种不平衡样本的分类预测问题;2)分类准确率高,总体预报准确率达到了 76.25% 以上,特别是稀有类样本的

分类正确率非常优秀,这是传统模型所无法做到的;3)对训练数据质量要求低,在高维数、数据缺失的情况下依然有较好的预测性能;4)无需事先对数据进行预处理以及特征选取;5)泛化性能好,避免过拟合;6)抗噪音,鲁棒性好.

4 结语

沙尘暴的短期发生与各气象指标存在非线性关系,并且具有一定的可预报性.然而,沙尘暴的发生属于稀有事件,由于稀有样本数量较少,而绝大部分样本都是属于不发生沙尘暴的平常样本,因此传统的预测模型都受到了极大干扰,几乎无法用于实际应用.基于这样的认识,借助统计学习理论,专门针对稀有类分类问题提出了沙尘暴短期预测模型,并取得了良好的预测能力.所建立的基于 SMOTE 算法与决策树算法的沙尘暴预测模型较好地解决了稀有类的分类问题,预测成功率超过了 75% .由于与传统模型比较,本模型在保持较高的总体预测精度的前提下,对于沙尘暴灾害的预报(稀有类)也达到了较高精度,因此可以用于实际的沙尘暴预警.显然,该项技术可以及时提醒沙源地附近居民及时做出反应,从而有助于将沙尘暴灾害的损失减少至最低.研究结果表明年平均风速、大风发生次数、年平均气温、年降水量、年平均相对湿度与沙尘暴天数均有较大的关联性,因此所选取的指标具有科学性,可对研究并预测沙尘暴的发生天数提供借鉴.

参考文献:

[1] 王金艳,王式功,马艳,等.我国北方春季沙尘暴与气候因子之关系[J].中国沙漠,2007,27(2):296-300.

[2] 夏训诚,杨根生.中国西北地区沙尘暴灾害及防治[M].北京:中国环境科学出版社,1996.

[3] 何清,赵景峰.塔里木盆地浮尘时空分布及对环境影响的研究[J].中国沙漠,1997,17(2):119-126.

[4] Littmann T.Dust storm frequency in Asia:climatic control and variability[J].International Journal of Climatology,1991(11):393-412.

[5] 杨东贞,房秀梅,李兴生.我国北方沙尘暴变化趋势的分析[J].应用气象学报,1998,9(3):352-358.

[6] 范一大,史培军,周俊华,等.近 50 年来中国沙尘暴变化趋势分析[J].自然灾害学报,2005,14(3):22-28

[7] McTains H G H,Lynch A W,Tews E K.Climatic controls upon dust storm occurrence in eastern Australia[J].Journal of Arid Environment,1998(9):457-466.

[8] McTains H G H,Burges S R C,Pitblado J R.A ridity drought and dust storms in Australia:1960—1984[J].Journal of Arid Environments,1989(16):11-22.

[9] McTains H G H,Lynch A W,Burgess R C.Wind erosion in eastern Australia[J].Journal of Soil Research,1990,28:323-339.

[10] 张小玲,李青春,谢璞,等.近年来北京沙尘天气特征及成因分析[J].中国沙漠,2005,25(3):417-521.

[11] 李岩瑛,李耀辉,罗晓玲,等.河西走廊东部沙尘暴预报方法研究[J].中国沙漠,2004,24(5):607-610.

[12] 黄富祥,张新时,徐永福.毛乌素沙地气候因素对沙尘暴频率影响作用的模拟研究[J].生态学报,2001,21(11):1875-1885.

[13] 张杰,郭铌,荻潇泓.西北地区东部冬季降水与次年沙尘暴发生的关系[J].中国沙漠,2004,24(5):603-606.

[14] 李锡福.青海省沙尘暴天气气候特征及其成因分析[C]//沙尘暴监测预警服务研究.北京:气象出版社,2002:227-232.

[15] 刘立超,安兴琴,李新荣,等.宁夏盐地沙尘暴特征分析[J].中国沙漠,2003,23(1):33-37.

[16] 赵建华,俞亚勋,孙国武.冻土对沙尘暴的影响研究[J].中国沙漠,2005,25(5):658-662.

[17] 赵光平,陈楠.生态退化状况下的宁夏沙尘暴发生发展规律特征[J].中国沙漠,2005,25(1):45-49.

[18] 郭铌,张杰,韩涛,等.西北特殊地形与沙尘暴发生的关系探讨[J].中国沙漠,2004,24(5):576-581.

[19] 矫梅燕,牛若芸,赵琳娜,等.沙尘天气影响因子的对比分析[J].中国沙漠,2004,24(6):696-700.

[20] 许炯心.黄土高原地区沙尘暴形成的自然地理因素:Ⅰ影响因素分析[J].中国沙漠,2005,25(4):547-551.

[21] 李登科,杜继稳.沙尘暴监测与预警方法研究[J].灾害学,2006,21(1):55-58.

[22] 钱正安,贺慧霞,瞿章,等.我国西北地区沙尘暴的分级标准和个例谱及其统计特征[C]//方宗义.中国沙尘暴研究.北京:气象出版社,1997.

[23] 胡金明,崔海亭,唐志尧.中国沙尘暴时空特征及人类活动对其发展趋势的影响[J].自然灾害学报,1999,8(4):49-56.

[24] Dong Z B,Wang X M,Liu L Y.Wind erosion in arid and semiarid China:an overview.[J].Journal of Desert Research,2000,20(2):134-139.

[25] 刘景涛,郑明倩.华北北部黑风暴的气候学特征[J].气象,1998,24(2):39-44.

[26] Chawla N V,Bowyer K W,Hall L O,et al.SMOTE:synthetic minority over-sampling technique[J].Journal of Artificial Intelligence Research,2002(16):321-357.
[27] 王存忠.中国沙尘暴"站时"时空变化特征及其与气象因子关系研究[D].南京:南京信息工程大学,2012.
[28] 叶笃正,丑纪范.关于我国华北沙尘天气的成因与治理对策[J].地理学报,2000,15(4):513-521.
[29] 张仁健,韩志伟,王明星,等.中国沙尘暴天气的新特征及成因分析[J].第四纪研究,2002,22(4):374-380.
[30] 李宁,顾孝天,刘雪琴.沙尘暴灾害致灾因子三维联合分布与重现期探索[J].地球科学进展,2013,28(4):490-496.
[31] 职为梅,范明.稀有类分类问题探讨[J].计算机技术与发展,2010,20(7):250-253.
[32] 谷振亚.稀有类分类算法在入侵检测中的应用[D].山西:太原理工大学,2010.

The Short-term Forecasting Model of Sandstorms Based on SMOTE and Decision Tree Learning Algorithm

ZHANG Zhenhua[a],XU Jinhui[b],LI Longxin[b],MA Chao[b],HUANG Jiangnan[a],XIE Zhenghong[c]

(Guangdong University of Foreign Studies a.School of Economics & Trade;
b.School of Finance; c.Cisco School of Informatics,Guangzhou 510006, China)

Abstract:As the traditional algorithms are defective in forecasting the accuracy of sandstorm disaster, this paper established a forecasting model combined SMOTE algorithm with decision tree learning algorithm.Using meteorological observation data of six provinces in Northwest China,the classification of rare class is well solved,and the predictive accuracy rate reaches 76.25%.The results showed that the model can be used for the actual sandstorm warning with good classification accuracy,generalization performance, robustness and anti-noise properties in solving the classification problems of the unbalanced samples in the sandstorm forecasting.

Key words:sandstorms forecasting; rare classes; SMOTE; decision tree learning algorithm

(编辑 徐永铭)