

熊亚军, 廖晓农, 李梓铭, 等. 2015. KNN 数据挖掘算法在北京地区霾等级预报中的应用. 气象, 41(1): 98-104.

KNN 数据挖掘算法在北京地区霾等级预报中的应用^{*}

熊亚军^{1,2} 廖晓农^{1,2} 李梓铭^{1,2} 张小玲^{1,2} 孙兆彬^{1,2}

赵秀娟^{1,2} 赵普生^{1,2} 马小会^{1,2} 蒲维维^{1,2}

1 京津冀环境气象预报预警中心, 北京 100089

2 中国气象局北京城市气象研究所, 北京 100089

提 要: 利用北京地区 2013 年气象数据以及 PM_{2.5} 浓度数据与能见度数据进行对比分析, 结果发现气温、气压、相对湿度、露点温度、地面 U 风、地面 V 风以及 PM_{2.5} 小时浓度这 7 个要素是影响北京地区霾等级的关键因素。利用气温、地面气压、相对湿度、露点温度、U 风、V 风分量以及 PM_{2.5} 浓度作为 7 个属性特征, 以霾等级做为标志量构建训练样本集, 结合 KNN (K-Nearest Neighbor) 数据挖掘算法构建霾等级预报分类器, 并开展霾等级客观识别实验。结果表明 K=3 时该分类器的分类预报效果最佳, 其 13 个站点的分类准确率高达 88.2%。基于该算法构建的 KNN 模型预报无霾时的漏报概率很小, 准确率高达 91.8%; 预报有轻度霾、中度霾以及重度霾时, 空报的概率仅分别为 4.7%、1.4% 和 2.6%。2014 年 8 月 29 日至 9 月 2 日北京地区一次霾天气过程的预报结果表明: 南郊观象台、密云和延庆 3 站的预报准确率分别达到 74%、64% 和 84%, 但霾等级的精度方面还有待于进一步提高。

关键词: 数据挖掘, KNN, 霾, 预报

中图分类号: P456, P413

文献标志码: A

doi: 10.7519/j.issn.1000-0526.2015.01.012

Application of KNN Data Mining Algorithm to Haze Grade Forecasting in Beijing

XIONG Yajun^{1,2} LIAO Xiaonong^{1,2} LI Ziming^{1,2} ZHANG Xiaoling^{1,2}

SUN Zhaobin^{1,2} ZHAO Xiujuan^{1,2} ZHAO Pusheng^{1,2} MA Xiaohui^{1,2} PU Weiwei^{1,2}

1 The Beijing-Tianjin-Hebei Environmental Meteorological Forecasting and Warning Center, Beijing 100089

2 Institute of Urban Meteorology, CMA, Beijing 100089

Abstract: Based on meteorological data and PM_{2.5} concentration data of 2013, the statistical analysis shows that temperature, pressure, relative humidity, dew point temperature, surface wind, U wind, V wind and PM_{2.5} hourly concentration are the 7 key factors affecting the haze grade of Beijing Area. The attribute vectors of KNN training sample set are the above 7 factors and the label vector is haze grade. The classifier can be established combining KNN Data Mining Algorithm and its best value is parameter K=3. The classification accuracy of the 13 meteorological stations is 88.2%. The forecasting model based on KNN Algorithm has good accuracy. When it predicts no haze, the accuracy rate is 91.8% and missing forecast chance is very small. When mild, moderate or severe haze is predicted, the probability of empty forecast is only 4.7%, 1.4% and 2.6%, respectively. There was one fog-haze weather process from 29 August to 2 September 2014 in Beijing Area. Its prediction results show the forecast accuracy of Nanjiao, Miyun and Yanqing is 74%, 64% and 84% respectively, but the accuracy of the haze grade remains to be further improved.

* 中央级公益性科研院所专项基金 (IUMKY201303PP0103)、国家科技支撑计划项目 (2014ABC16B04)、北京市科技计划项目 (Z131100006113013) 和首都蓝天行动培育专项 (Z141100001014013) 共同资助

2014 年 9 月 4 日收稿; 2014 年 11 月 18 日收修定稿

第一作者: 熊亚军, 主要从事环境气象预报预警及决策服务工作. Email: xyajun78@163.com

Key words: data mining, K-Nearest Neighbor (KNN), haze, forecast

引言

霾是大量极细微的尘粒等均匀地浮游在空中,使水平能见度 <10 km 的空气普遍混浊现象(中国气象局,2010)。然而,随着社会经济以及城市化进程的不断加快,能源消耗和人类活动导致我国京津冀、长三角、珠三角和成渝等大型城市群的空气中细颗粒物浓度明显增加。空气质量变差不仅对市民的健康构成了威胁,而且由于气溶胶粒子的消光作用会影响能见度,特别是它们吸湿增长后使能见度大幅度下降,从而给交通出行带来很大影响(段玉森等,2005;吴兑等,2011;吴兑,2012)。例如,2013 年 1 月和 2014 年 2 月末京津冀地区伴随严重污染的持续性低能见度天气就导致航班大面积延误、高速公路长时间封闭。因此,当今霾已经不是一种纯自然现象,霾的等级预报也已经成为很多气象局的常规业务。

大量研究表明,霾的发生与地面风和相对湿度等气象要素存在密切的关系。当前国内外的霾预报主要有数值预报和模式输出统计预报(model output statistics, MOS)两大类方法。数值预报方法是在大气动力学模式的基础上耦合了化学模块,直接模拟大气污染与大气气象要素场的变化过程(邓涛等,2012;2013;沈劲等,2011)。在业务化过程中,这类方法受到排放清单不确定等因素影响,在日常业务化预报中有较大的局限。MOS 方法主要思路是:首先依据污染物浓度数据集筛选出关联性强的气象因子,然后利用统计方法建立预报方程,最后利用模式输出量代入预报方程进行统计预报。陈亦君等(2014)应用基于系统辨识理论的实时迭代模式对 WRF 模式预报结果进行后处理,建立了上海地区霾天气的模式输出-统计预报系统,结果发现霾日预报成功率达到 72.7%~73.7%。程一帆等(2011)在分析能见度与空气污染浓度关系的基础上,直接利用污染扩散方程建立兰州市冬季霾统计预报模型,在实际预报业务取得较好预报效果。毛宇清等(2011)基于支持向量机(support vector machine, SVM)建立南京地区霾日分类预报模型,结果表明 Ts 评分均在 0.4 以上,有霾日 14 时能见度 SVM 回归预报结果的准确率在 86% 以上。郑峰等(2011)

利用温州 8 个气象站点近 40 年雾日和霾日统计资料,建立包含晴雨、空气污染指数及雾、霾 3 要素的线性预报系统,发现未来 24 h 雾、霾预报准确率达 57%~66%。上述处理方法均单一地采用了数值天气模式的结果,主要基于霾与气象条件之间的相关性(过宇飞等,2013;刘勇洪等,2014)。如前所述,在污染日益严重的背景下,霾的严重程度受到气溶胶浓度的影响很大。但是,目前尚没有同时采用数值天气预报模式和环境模式的两类输出结果的相关试验研究。

鉴于霾预报的复杂性,本文尝试一种新的思路:即同时选用气象和环境两类不同模式的输出结果(分别为 EC 模式输出的气象场和 BREMPMS 模式输出的 $PM_{2.5}$ 浓度预报场),再选用恰当的算法对两类数据进行融合,通过建立动态的预报模型实现霾的等级预报。近年来一些数据挖掘算法近年来在气象领域受到关注。KNN 算法作为十大数据挖掘的十大经典算法之一,在风(邵明轩等,2006;陈豫英等,2008;涂小萍等,2008)、降水预报(曾晓青等,2008)、云分类(朱彪等,2012)均有应用。本文尝试将 KNN 算法应用到北京地区霾等级分类中,以期建立客观化的霾等级预报系统提供有力的科技支撑。

1 资料和方法

1.1 资料来源

本文所用的常规气象观测资料来源于北京市信息中心,包括温度、气压、湿度、风和能见度等气象要素。环境数据来源于环保部在其官方网站(<http://113.108.142.147:20035/emcpublish>)公布的 $PM_{2.5}$ 小时浓度资料。资料年限为 2013 年 1 月 1 日至 12 月 31 日。构建训练样本集时,取北京地区 13 个区(县)因缺少对应的环保局 $PM_{2.5}$ 数据未录入房山区数据集。一天 8 次、间隔 3 h 的地面气象资料和 $PM_{2.5}$ 数据。

1.2 KNN 算法

随着计算机技术的逐渐发展,数据挖掘技术目前在气象中的应用越来越多。黄颖等(2014)将局部线性嵌入算法应用到台风强度集合预报模型,发现

能够提高预报效果。刘旸等(2011)利用神经网络法反演晴空大气湿度廓线的研究,发现所获得的水汽廓线更加贴近真实廓线。农孟松等(2011)应用主分量分析(PCA)方法,对同一降水预报量的各种数值预报产品进行集成预报研究,发现新的集成预报模型不仅对历史样本的拟合精度好于各个子预报产品,而且对独立样本的实验预报结果也显示出更好的预报准确率及稳定性。KNN 是 Cover 等(1967)提出的一种非参数化监督算法,也是十大经典数据挖掘技术之一,其在统计学领域通常译作“K 近邻算法”。KNN 是一种通过计算对象间不同特征值距离方法来进行分类的非参数估计技术,也是近年来发展较快的一种实用数据挖掘技术。该算法的优点是精度高、对异常值不敏感、不需要数据输入假定,其缺点是计算复杂度、内存空间开销较大,但随着计算机硬件技术的不断发展,这些缺点变得越来越容易解决。

虽然建立回归预报方程预报法仍然是当前数值天气预报解释应用的一种基本方法,但由于大气运动本身具有混沌性和非线性特征,因此事实上很难用一个或一组简单的线性或非线性回归方程对天气模型精确描述和模拟,故回归预报技术的实际应用效果有时差强人意。KNN 算法的核心思想是:客观世界具有规律性和重现性,在相似的条件往往产生相似的结果,该算法在云分类、风力预报和医疗气象中均有应用。KNN 算法在数值天气预报解释应用时,通常直接基于历史天气个例样本建立训练和预测模型,并将天气学预报思路和数值预报结果进行某种融合,故在某些情形下,KNN 发挥算法本身的优势,避开建立固定的回归预报方程等不利因素影响。

具体步骤如下:假设有一组历史天气个例样本集合为定义为 S 。其中, S 由 i 个天气样本组成,而每个天气样本由 m 个属性变量以及 1 个标志量构成。其数学表示式为:

$$S = \left\{ \begin{array}{l} [X_{11}, X_{12}, X_{13}, \dots, X_{1m}, L_1] \\ [X_{21}, X_{22}, X_{23}, \dots, X_{2m}, L_2] \\ \dots\dots\dots \\ [X_{i1}, X_{i2}, X_{i3}, \dots, X_{im}, L_i] \end{array} \right\}$$

本文的属性变量 X_{ij} 分别代表温度、地面气压、相对湿度、露点温度、地面 U 风、地面 V 风以及 $PM_{2.5}$ 浓度这 7 个要素,最后一个要素 L_i 称为标志

量(label),本文中的标志量即为霾等级。利用 KNN 算法进行霾等级预测可以描述为如下数学模型:假设预报日的温度、气压、相对湿度、露点温度、地面 U 风、地面 V 风以及 $PM_{2.5}$ 浓度 7 个要素的集合为 $Y_i = \{y_1, y_2, \dots, y_m\}$,称为预测样本;预测时,首先在训练样本集 S 集合中找到与预测样本 Y_i 最相似的 K 个近邻(K 通常为奇数);然后找出这 K 个标志量(即霾等级)集合 $L = \{L_1, L_2, \dots, L_k\}$;最后,按照投票多数原则,选取最多的标志量 L_i 作为预测样本 Y_i 预测结果。本文的 K 近邻居选取采用欧氏距离法进行判定。因此,从本质上说,KNN 算法是一种相似算法,KNN 算法还被广泛应用于手写码识别(姜文等,2013)和人脸识别(杨淑平等,2013)等领域。本文针对北京的 13 个气象站点分别构建各自的 KNN 分类器进行预测。预测时,本文所用的气象要素基于 ECMWF 细网格模式获得, $PM_{2.5}$ 浓度数据基于 BREMPs 模式获得。

2 试验结果及分析

2.1 KNN 分类器中属性变量选取

基于 KNN 数据挖掘算法进行霾等级预报的关键是构建合理的 KNN 分类器,而属性变量选取合适与否对分类器的准确率至关重要的影响。本文的 KNN 分类器属性变量基于相关分析获得。由霾的定义可知,相对湿度和能见度是定义霾等级的两个关键要素。表 1 给出了它们与能见度的相关系数。图 1 给出了相关气象要素与能见度的散点图。由图 1a 及表 1 可知, $PM_{2.5}$ 与能见度的相关系数为 0.69,通过了 0.001 的显著性水平检验。两者可用关系式 $y = -8.337 \ln x + 49.726$ 来进行拟合(式中 y 为能见度, x 为 $PM_{2.5}$ 浓度)。由图 1a 还可以发现,当 $PM_{2.5}$ 浓度 $< 70 \mu g \cdot m^{-3}$ 时,随着浓度增大,能见度急剧变小;反之,当 $PM_{2.5}$ 浓度 $> 70 \mu g \cdot m^{-3}$ 时,随着浓度增高,其对能见度的影响程度逐渐减小。由图 1b 可知,露点温度与能见度之间呈现负相关关系,其相关系数 $r = -0.31$,通过 0.01 的显著性水平检验(表 1),即随着露点温度升高(水汽含量增加),能见度逐渐减小。图 1c 给出相对湿度和能见度的相关关系。相对湿度与能见度之间也呈现负相关关系,相关系数为 -0.49 ,通过 0.001 显著性水平检验。能见度和相对湿度之间可用关系式 $y =$

-11.22lnx + 59.767 来进行拟合(式中 y 为能见度, x 为相对湿度),即随着相对湿度增加,能见度逐渐减小。由图 1c 进一步可以发现,当相对湿度 < 60%时,随着相对湿度的增加,能见度减小的程度较大;反之,当相对湿度 > 80%时,随着相对湿度的增加,能见度减小的程度较小。王京丽等(2006)发现空气中的细粒子浓度与大气能见度存在显著的相关关系;曹伟华等(2013)在分析北京的低能见度天气时,也发现相对湿度和 PM_{2.5} 浓度是影响能见度大小的两个关键因子。图 1d 则反映了风速与能见度之间的相关关系。风速与能见度的相关系数为 0.23,两者之间整体呈现正相关关系,即随着风速增大,扩散条件转好,能见度逐渐增大。同时,图 1d 也表明当风速 < 2 m · s⁻¹ 时(即传统意义上的小风速),能见度一般 < 10 km;但是当风速 > 2 m · s⁻¹ 且 < 4 m · s⁻¹ 时,既有很大一部分样本的能见度在 10 km 以上。

上述要素虽然都与能见度有一定的影响,但用

单一的气象要素肯定不能很好表征能见度的变化,这也充分说明了能见度变化的复杂性。以风为例,在实际业务值班过程中经常发现,当华北区域整体盛行 2~4 m · s⁻¹ 的偏北风时(其通常意味着弱冷空气南下,常见于霾天气的结束期),由于扩散条件转好,北京地区的能见度逐渐转好;而当华北南部地面盛行 2~4 m · s⁻¹ 的偏南风时,由于其将河北南部的气溶胶例子输送到北京地区,并导致北京地区 PM_{2.5} 浓度升高,因此,通常导致本地能见度的下降。但是,当南风风速过大时(5~8 m · s⁻¹)时,虽然气溶胶输送也很明显,过大的地面风也意味着低层扩散条件转好,能见度反而有所降低。风速、风向对能见度均有较大的影响,这也是 KNN 训练集中将风分解为 u 、 v 风两个分量的原因。

另外,鉴于气温和气压是表征天气统的两个最基本的要素,其对冷空气活动、天气系统的表征具有关键意义,因此也入选了训练属性集。

表 1 各种气象要素与能见度的相关系数

Table 1 Correlation coefficients between meteorological elements and visibility value

气象要素	相对湿度	露点温度	风速	PM _{2.5} 浓度
与能见度的相关系数	-0.49	-0.31	0.23	-0.69

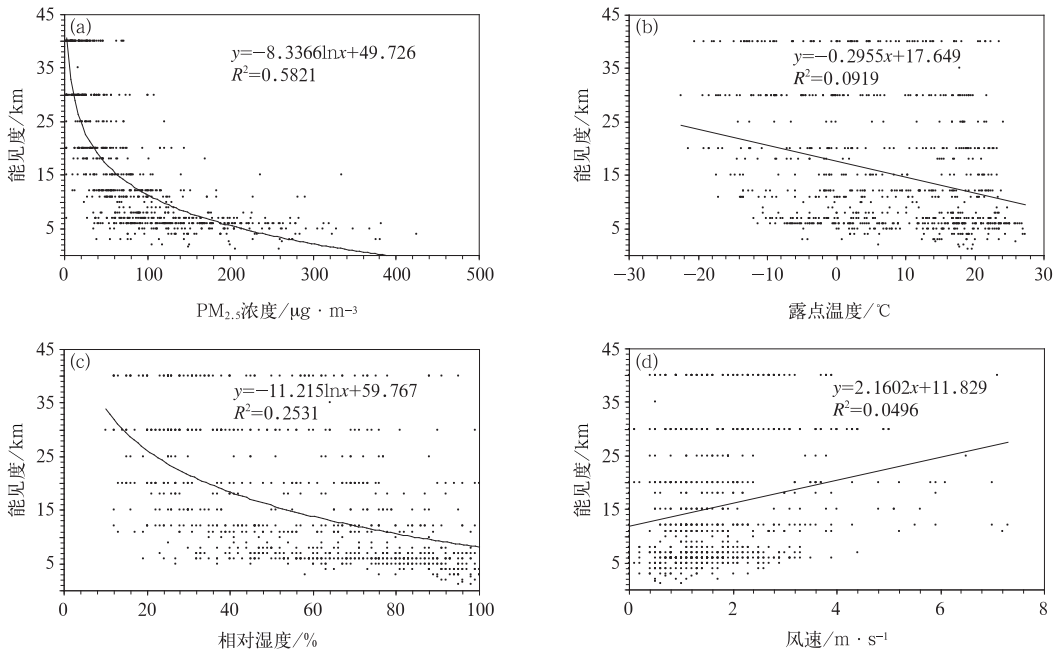


图 1 能见度与各种气象要素的散点图

(a) PM_{2.5} 浓度, (b) 露点温度, (c) 相对湿度, (d) 风速

Fig. 1 scatter plots between visibility and various meteorological elements

(a) concentration of PM_{2.5}, (b) Dew point temperature,

(c) relative humidity, (d) wind speed

2.2 K 参数的选取

KNN 算法中 K 值的选取对于分类器的准确率有较大的影响。为便于在最邻近的子集中进行投票, K 值的选取一般为奇数。若 K 值太小, 则 KNN 分类结果容易受到噪声的影响; 反之, 若 K 的取值过大, 则分类器提取的最邻近子集中容易因为包含过多的错误点而对分类准确率产生影响。

交叉验证 (Cross validation) 俗称循环估计法, 其统计学上将数据样本切割成较小子集进行验证的实用方法。通常先在一个子集上做分析, 而其他子集则用来做后续对此分析的确认及验证。交叉验证是一种评估统计分析、机器学习算法对独立于训练数据的数据集的泛化能力的实验方法。本文的 K 值采用交叉验证方式进行确定。表 2 给出了 K 分别取 3、5 和 7 时的分类准确率。结果表明 (表 2): 当 $K=3$ 、5 或 7 时, 北京各气象站的交叉检验准确率均在 80.0% 以上, 13 个站点的平均准确率分别为 88.2%、85.8% 和 84.7%, 且 $K=3$ 时, 各站的分类

表 2 K 取不同值时 KNN 算法的交叉检验准确率 (单位: %)

Table 2 KNN algorithm accuracy through Cross validation when taking different values (unit: %)

站点	$K=3$	$K=5$	$K=7$
顺义	88.4	87.3	86.7
海淀	87.0	83.6	82.4
延庆	89.0	86.3	84.8
密云	87.3	85.0	83.6
怀柔	85.1	81.0	80.0
平谷	90.4	88.0	87.1
通州	88.9	87.5	87.7
朝阳	87.7	84.4	83.5
昌平	90.5	88.1	87.1
门头沟	91.4	89.6	88.0
南郊	85.5	81.3	79.4
石景山	88.3	86.6	85.4
丰台	86.7	86.7	85.1
平均准确率	88.2	85.8	84.7

准确率明显优于 $K=5$ 或 7 时的结果, 而 $K=5$ 的结果又优于 $K=7$ 的结果。因此, 本文采用 $K=3$ 来建立 KNN 模型。

2.3 KNN 分类器的准确率分析

为了解 KNN 分类器对各级霾的预报性能, 表 3 进一步给出了 KNN 分类器对各级霾分类的交叉检验结果。表 3 中横轴代表霾的预报, 横轴表示霾的实况, 交叉线 (底色为红色的表格) 代表预报准确率。由表 3 可知: (1) 当 KNN 分类器预报没有霾时, 预报准确率高达 91.8%; 轻微霾漏报的概率为 6.9%, 而轻度霾、中度霾和重度霾漏报的概率均 $< 1\%$ 。(2) 当 KNN 分类器预报有轻微霾时, 空报的概率为 16.1%, 预报正确的概率高达 67.4%, 轻度霾漏报的概率为 11.2%, 中度和重度霾漏报的概率均 $< 5\%$ 。(3) 当 KNN 分类器预报有轻度霾时, 空报的概率为 4.7%, 预报正确的概率高达 59.8%。实况较预报偏轻, 为轻微霾的概率为 21.4%; 实况较预报偏重, 为中度和重度霾漏报的概率分别为 10.4% 和 3.7%。(4) 当 KNN 分类器预报有中度霾时, 空报的概率仅有 1.4%, 预报正确的概率为 53.4%; 实况较预报偏轻, 为轻微霾和或轻度霾的概率分别为 13.9% 和 18.8%; 实况较预报偏重 (实况未重度霾) 的概率为 12.5%。(5) 当 KNN 分类器预报有重度霾时, 空报的概率为 2.6%, 等级预报完全正确的概率为 60.4%。

综上所述, KNN 分类器预报霾的有无时空报或漏报率较低, 而且霾的等级预报的准确率较高。但是, 当模型预报有霾时, 对于相邻等级的区分存在一定的误差。事实上, 按照 QX/T 113-2010 标准进行霾的等级划分时, 有些等级范围较窄 (例如: 规定中度霾能见度为 2~3 km), 加上能见度观测本身存在一定的误差, 因此本文构建的 KNN 分类器具有一定的业务实用价值。

表 3 KNN 分类器对各级霾的交叉检验结果 (单位: %)

Table 3 Results of KNN classifier for all levels of haze through cross validation (unit: %)

实况预报	无霾	轻微霾	轻度霾	中度霾	重度霾
无霾	91.8	16.1	4.7	1.4	2.6
轻微霾	6.9	67.4	21.4	13.9	16.2
轻度霾	0.8	11.2	59.8	18.8	8.4
中度霾	0.3	3.6	10.4	53.4	12.3
重度霾	0.1	1.8	3.7	12.5	60.4

3 基于 KNN 算法的实际业务预报系统

本研究所用的 KNN 算法由 Python 语言进行编码。由于 Python 语言是一种脚本语言,因此,很便于气象领域搭建实际的业务预报系统。笔者基于 KNN 算法,搭建了北京地区站点霾等级预报系统。平台的气象要素预报由 EC 细网格模式输出场通过插值获得,PM_{2.5} 浓度则由 BREMPs 模式获得。平台每天 20 时启动一次,起报未来 72 h 的霾等级,时间分辨率为 3 h。2014 年 8 月 29 日至 9 月 2 日,北京地区出现了一次霾天气过程,部分站点伴随降水。图 2 给出了北京地区 3 个基本站(南郊、密云和延庆)72 h 时效内间隔 3 h 的预报与实况的对比。图 2a

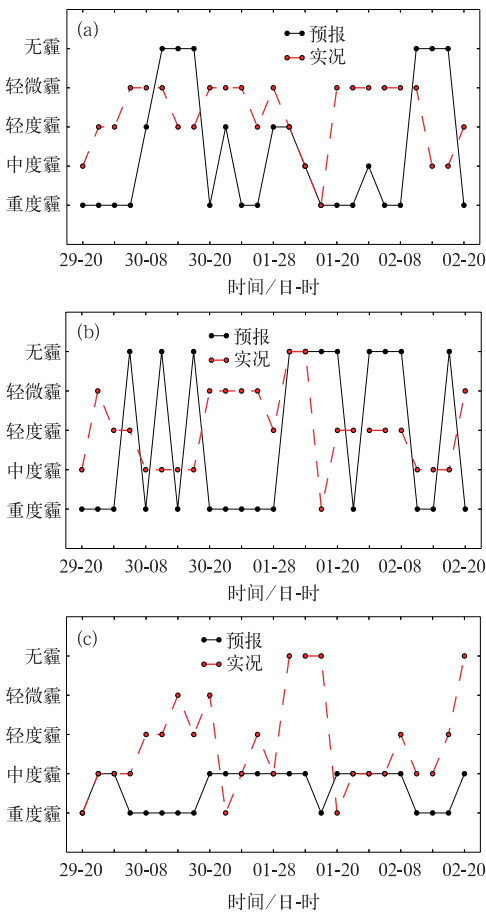


图 2 基于 KNN 算法制作的北京地区霾等级预报业务产品

(a)南郊观象台,(b)密云站,(c)延庆站

Fig. 2 Haze grade forecast products made based on KNN algorithm in Beijing Area

(a)Nanjiao Observatory, (b)Miyun Station, (c)Yanqing Station

表明,所有 25 个观测时次均出现了不同程度的霾,而预报则有 19 个时次有霾,霾的有无准确率达到 76%。但是在 8 月 29 日 08 时的初始阶段,实况没有轻微一中度霾,KNN 预报系统预报偏重。8 月 29 日 14 时和 9 月 2 日 14 时,实况为轻度一中度霾。KNN 预报系统预报为无霾,预报偏轻。经过天气分析可知,8 月 29 日 14 时和 9 月 2 日 14 时 EC 细网格模式的相对湿度预报较实况偏低,这可能也是 KNN 模式预报偏轻的主要因素。密云站(图 2b)和延庆站(图 2c)的有无预报准确率分别达到 64%和 84%,但在霾的等级精确程度略有欠缺。

4 结 论

本文研究了将基于数据挖掘技术的 KNN 算法与天气模式和大气成分模式相结合实现霾的分级客观预报的方法,得到以下结论:

(1)霾预报的 KNN 分类器属性特征由温度、气压、相对湿度、露点温度、U 风、V 风以及 PM_{2.5} 浓度等 7 变量构成,标志项为霾的等级;

(2)KNN 分类器分别 K 取 3、5 或 7 的实验结果表明,北京地区 13 个站点的交叉检验准确率分别为 88.2%、85.8% 和 84.7%,且 $K=3$ 时的分类准确率明显优于 $K=5$ 或 7 时的结果。当 KNN 模型预报的整体预报准确率较高。预报无霾时的准确率高达 91.8%,漏报概率很小。KNN 模型预报有霾时,空报的概率较低,预报准确率较高;

(3)2014 年 8 月 29 日至 9 月 2 日北京地区一次霾天气过程的预报结果表明,南郊观象台、密云和延庆站的霾有无预报准确率分别达到 74%、64% 和 84%。

应该指出,由于受到大气成分监测资料的限制,本文只采用了 2013 年的分析数据作为训练样本集,在一定程度上影响了预报准确率。而且,目前 KNN 霾分类算法直接运用了大气化学模式 BREMPs 的结果,接下来待 PM_{2.5} 浓度资料累积到一定程度后,可对 BREMPs 的预报结果进行必要的订正。因此,KNN 霾分类算法的准确率仍然有进一步提升的空间。

参考文献

曹伟华,梁旭东,李青春. 2013. 北京一次持续性雾霾过程的阶段性特征及影响因子分析. 气象学报,71(5):940-951.

- 陈亦君, 尤佳红, 束炯, 等. 2014. 基于 WRF-RTIM 的上海地区霾预报 MOS 方法研究. 环境科学学报, 34(3):574-581.
- 陈豫英, 刘还珠, 陈楠, 等. 2008. 基于聚类天气分型的 KNN 方法在风预报中的应用. 应用气象学报, 19(5):564-572.
- 程一帆, 张兰慧, 尚可政, 等. 2011. 兰州市冬季霾的统计预报研究. 中国沙漠, 31(3):750-756.
- 邓涛, 吴兑, 邓雪娇, 等. 2012. 珠江三角洲一次典型复合型污染过程的模拟研究. 中国环境科学, 32(2):193-199.
- 邓涛, 吴兑, 邓雪娇, 等. 2013. 珠三角空气质量暨光化学烟雾数值预报系统. 环境科学与技术, 36(4):62-68.
- 段玉森, 束炯, 张弛, 等. 2005. 上海市大气能见度指数指标体系的研究. 中国环境科学, 25(4):460-464.
- 过宇飞, 刘端阳, 周彬, 等. 2013. 无锡市霾天气特征及影响因子研究. 气象, 39(10):1314-1324.
- 黄颖, 金龙, 黄小燕, 等. 2014. 基于局部线性嵌入的人工智能台风强度集合预报模型. 气象, 40(7):806-815.
- 姜文, 卢朝阳, 李静. 2013. 基于方向线索特征的手写体维文字符识别. 微电子学与计算机, 30(10):97-100.
- 刘畅, 官莉. 2011. 人工神经网络法反演晴空大气湿度廓线的研究. 气象, 37(3):318-324.
- 刘勇洪. 2014. 基于 NOAA/AVHRR 卫星资料的北京地区霾识别研究. 气象, 40(5):619-527.
- 毛宇清, 孙燕, 姜爱军, 等. 2011. 南京地区霾预报方法试验研究. 气候与环境研究, 16(3):273-279.
- 农孟松, 黄海洪, 孙崇智, 等. 2011. 基于主分量神经网络的降水集成预报方法研究. 气象, 37(3):352-355.
- 邵明轩, 刘还珠, 窦以文. 2006. 用非参数估计技术预报风的研究. 应用气象学报. 17(S1):125-129.
- 沈劲, 王雪松, 李金凤, 等. 2011. Models-3 /CMAQ 和 CAMX 对珠江三角洲臭氧污染模拟的比较分析. 中国科学(化学), 41(11):1750-1762.
- 涂小萍, 赵声蓉, 曾晓青, 等. 2008. KNN 方法在 11—3 月中国近海测站日最大风速预报中的应用. 气象, 34(6):67-73.
- 王京丽, 刘旭林. 2006. 北京市大气细粒子质量浓度与能见度定量关系初探. 气象学报, 64(2):221-228.
- 吴兑. 2012. 近十年中国灰霾天气研究综述. 环境科学学报, 32(2):257-269.
- 吴兑, 吴晟, 李海燕, 等. 2011. 穗港晴沙两重天——2010 年 3 月 17—23 日珠三角典型灰霾过程分析. 环境科学学报, 31(4):695-703.
- 杨淑平, 易国栋, 袁修贵, 等. 2013. 一种基于分块小波的人脸识别算法. 中南大学学报(自然科学版), 44(5):1902-1909.
- 郑峰, 颜琼丹, 吴贤笃, 等. 2011. 温州地区雾霾气候特征及其预报. 气象科技, 39(6):791-795.
- 曾晓青, 邵明轩, 王式功, 等. 2008. 基于交叉验证技术的 KNN 方法在降水预报中的试验. 应用气象学报, 19(4):471-478.
- 中国气象局. 2010. 中华人民共和国气象行业标准 QX/T113—2010, 霾的观测和预报等级. 北京:气象出版社.
- 朱彪, 杨俊, 吕伟涛, 等. 2012. 基于 KNN 的地基可见光云图分类方法. 应用气象学报, 23(6):721-728.
- Cover T M, Hart P E. 1967. Nearest neighbor pattern classification. IEEE Trans Inform Theory, 13(1):21-27.