

Automatic Identification of Two Kinds of Weather Image Based on Bayesian Classification

Hao Yu*, Xiaotong Wang, Guanlei Xu

Dalian Naval Academy, Dalian Liaoning
Email: *995034849@qq.com

Received: Oct. 3rd, 2015; accepted: Oct. 17th, 2015; published: Oct. 21st, 2015

Copyright © 2015 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

To realize the automatic observation of weather phenomena, a method for recognizing weather phenomena of outdoor image based on Bayesian classification is proposed. This method extracts image characteristics of hue, saturation and brightness, and adopts Bayesian classification to estimate the haze and sand-dust weather. The accuracy of experiment results reaches the expected aim, and the experimental results show that this method can quickly and accurately measure the weather phenomena of outdoor image.

Keywords

Weather Phenomena Identification, Image Processing, Bayesian Classification

基于贝叶斯分类的两类天气图像自动识别

于浩*, 王孝通, 徐冠雷

海军大连舰艇学院, 辽宁 大连
Email: *995034849@qq.com

收稿日期: 2015年10月3日; 录用日期: 2015年10月17日; 发布日期: 2015年10月21日

*通讯作者。

摘要

为实现天气现象的自动观测，本文提出了一种基于贝叶斯分类的室外图像的天气现象识别方法，该方法通过提取图像的色相、饱和度和亮度等特征，采用贝叶斯分类，对雾霾和沙尘两类天气现象进行判断。实验结果的准确率达到预期目标，并且能快速、准确地测量室外图像的天气现象。

关键词

天气现象识别，图像处理，贝叶斯分类

1. 研究目的和现状

最近随着图像理解与智能视频技术的不断发展，对室外环境下各种天气现象，尤其是恶劣天气现象的自动检测与识别提供了必要条件。但是，目前天气现象的观测仍然主要依靠人工观测，天气现象信息的采集耗费了大量的人力和物力，而且人工观测受外界环境和地域条件影响较大，观测数据较少且不连续，因此近年来基于室外图像与视频数据的天气现象自动识别研究受到了较为广泛的关注[1]。

目前，人们在不同方向对各种天气现象进行分析并建模，取得了一定的成果。本文提出了一种通过提取的图像特征进行贝叶斯分类的天气现象识别与分类方法。该方法通过提取图像的色相、饱和度和亮度等特征，采用贝叶斯分类，对雾霾和沙尘两类天气现象进行判断。该方法可以进行连续观测，且能快速准确的识别天气现象[2]。

2. 贝叶斯理论和方法

Pearl 通过结合专家系统，发展出贝叶斯理论的高级应用模型贝叶斯网络，并指出贝叶斯网络或许是概率推理中最普及的模型[3]。贝叶斯理论经过长期的发展与运用，其理论被很多学者运用到各行各业，是统计学中一支重要的理论。

贝叶斯方法是基于贝叶斯定理而发展起来用于系统地阐述和解决统计问题的方法。该方法的核心为贝叶斯公式[4]，其基本形式如下：

$$P(w|x) = \frac{P(x|w)p(w)}{P(x)}$$

式中， $P(w)$ 代表对 w 拥有的初始概率，即 w 的先验概率，它反映了关于 w 是一正确假设的机会有背景知识。 $P(x)$ 代表将要观察的集合 x 的全概率，即在没有确定某一假设成立时 x 的概率。 $P(x|w)$ 代表假设 w 成立的情形下观察到集合 x 的概率，即条件概率。 $P(w|x)$ 代表给定集合 x 时 w 成立的概率，即 w 的后验概率，它反映了在看到集合 x 后 w 成立的置信度。

贝叶斯理论经过长期的历史发展，以其独特的不确定性知识表达形式、丰富的概率表达能力、综合先验知识的增量学习特性等成为众多学者日益研究的焦点问题。贝叶斯在现实生活中作为数据挖掘、分类等方面得到了广泛的应用。其中尤其发展而来的朴素贝叶斯分类器得到了广大学者的重视，朴素贝叶斯分类器(naive Bayes classifiers, 简记为 NB)基于简单而不现实的假设：在给定类标记时属性值之间相互条件独立[5]。

贝叶斯定理作为一种概率计算可用于多个领域内进行概率推理。今天，我们用贝叶斯法则过滤垃圾

邮件，为网站用户推荐唱片、电影和书籍。它渗透到了互联网、语言和语言处理、人工智能、机器学习、金融、天文学和物理学乃至国家安全等各个领域，当然也可以运用在图像识别概率的优化上，从而建立数学优化模型[6]。

3. 朴素贝叶斯模型的建立

在许多场合下，我们研究贝叶斯的重点落在修订其后验概率的工作上。为了寻找给定集合 x 时可能性最大的假设 $w \in W$ (W 为候选假设集合)。这种具有最大可能性的假设被称为极大后验假设，简记为 w_{map} 。即

$$w_{map} = \arg \cdot \max P(w_n | x_m) \quad (3.1)$$

应用到贝叶斯公式得到：

$$w_{map} = \arg \cdot \max \frac{P(x_m | w_n) P(w_n)}{P(x_m)} \quad (3.2)$$

因为 $P(x_m)$ 是一个不依赖于 w_n 的常量，去掉它可得到：

$$w_{map} = \arg \cdot \max P(x_m | w_n) P(w_n) \quad (3.3)$$

为做好几种天气图像的识别，我们需要将实际问题与贝叶斯原理进行紧密的结合。实际上，图像识别通常与饱和度、色相等要素有着密切的关联，即给定一组因子，对因子进行合理的分类和数据挖掘，就可以得到显著的相关关系，从而进行概率的优化与完善。我们可以将天气图像识别与朴素贝叶斯分类器结构相结合，以自变量表示图像的饱和度等要素，以因变量表示研究对象及图像所属的天气种类。显然，分类问题就可以直接用 MAP 假设来解决[7]。

应用 MAP 假设分类， x 的目标是在给定集合 $\langle x_1, x_2, x_3, \dots, x_m \rangle$ 的情况下，得到最可能的 $w(x)$ 的值。应用公式(3.3)得到

$$w(x)_{map} = \arg \cdot \max P(x_1, x_2, x_3, \dots, x_m | w_n) P(w_n) \quad (3.4)$$

现在要做的就是基于总集合估计上式中的两个概率值。估计每个 $P(w_n)$ 值很容易，只要计算每个 w_n 出现在集合 W 中的频率就可以。然而，估计每个 $P(x_1, x_2, x_3, \dots, x_m | w_n)$ 的值不太可行，原因在于：首先，完整估计 $P(x_1, x_2, x_3, \dots, x_m | w_n)$ 值的时候复杂程度相当于学习一个贝叶斯网络，是一个比较难的问题。因此，为获得合理的估计，我们引入朴素贝叶斯分类器机构，假定：在给定自变量时其属性值之间是相互条件独立的。也就是说，在给定自变量集合的情况下，观察到的联合概率正好是每个单独属性值概率的乘积。

具体数学表达式如下：

$$P(x_1, x_2, x_3, \dots, x_m | w_n) = \prod_{j=1}^m P(x_j | w_n) \quad (3.5)$$

将其带入公式(3.4)可得到朴素贝叶斯分类器的分类公式：

$$w(x)_{map} = \arg \cdot \max P(w_n) \prod_{j=1}^m P(x_j | w_n) \quad (3.6)$$

式中， x_j 为预报因子 x 的第 j 个属性、概率 $P(w_n)$ 和 $P(x_j | w_n)$ 可以通过计算集合中不同种类和属性值组合的出现频率来简单计算，计算公式如下：

$$P(w_n) = \frac{\sum_{i=1}^n \delta(w_i, w_n)}{n} \quad (3.7)$$

$$P(x_j | w_n) = \frac{\sum_{i=1}^n \delta(x_{ij}, x_j) \delta(w_i, w_n)}{\sum_{i=1}^n \delta(w_i, w_n)} \quad (3.8)$$

式中, n 为集合的总个数, w_i 为集合中第 i 个要素, x_{ij} 为第 i 个预报因子的第 j 个属性, $\delta(w_i, w_n)$ 为一个二值函数, 当 $w_i = w_n$ 时为 1, 否则为 0。

在大多数情况下, 上述这种基于频率比例的方法是对概率的一个良好估计, 但当接近零频率属性值出现的时候, 就会导致这种概率估计产生一个有偏的过低估计概率。更极端的是, 当零频率属性值出现时, 就会使得某概率值为 0, 进而致使式(3.6)为 0。为了避免此类问题的出现, Laplace 估计常常被用来平滑上述所得概率, 修改式(3.7)和(3.8)可得:

$$P(w_n) = \frac{\sum_{i=1}^n \delta(w_i, w_n) + 1}{n + n_w} \quad (3.9)$$

$$P(x_j | w_n) = \frac{\sum_{i=1}^n \delta(x_{ij}, x_j) \delta(w_i, w_n) + 1}{\sum_{i=1}^n \delta(w_i, w_n) + n_j} \quad (3.10)$$

式中, n_w 为因变量种类的个数, n_j 为该预报因子属性的种类个数。

4. 朴素贝叶斯模型的应用

运用朴素贝叶斯分类器来解决天气图像识别问题: 根据图像的要素来判断图像中的天气属于哪一类 [8]。从试验数据中筛选一组数据集合(部分内容见表 1), 我们可以将因变量划分为两类:

雾霾天气 W_N 和沙尘天气 W_Q

Table 1. Part of the test data set

表 1. 试验数据集合(部分)

序号	色相均值	饱和度均值	亮度均值	天气现象
1	0.3008	0.0067	0.7058	雾霾
2	0.5671	0.0693	0.5632	雾霾
3	0.5402	0.1159	0.6693	雾霾
4	0.5867	0.3455	0.5803	雾霾
5	0.4532	0.0452	0.5539	雾霾
6	0.1084	0.6269	0.7579	沙尘
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
$n-2$	0.0973	0.5314	0.7281	沙尘
$n-1$	0.0878	0.4673	0.5568	沙尘
n	0.0946	0.5699	0.6213	沙尘

可以从原始数据中选取合适的自变量：色相均值 X_1 、饱和度均值 X_2 、亮度均值 X_3 。其中每一类自变量我们可以将其划分为三种属性[9]：

$$\text{色相均值 } X_1: X_{11} \in [0, 0.3]; X_{12} \in [0.3, 0.6]; X_{13} \in [0.6, 1]$$

$$\text{饱和度均值 } X_2: X_{21} \in [0, 0.3]; X_{22} \in [0.3, 0.6]; X_{23} \in [0.6, 1]$$

$$\text{亮度均值 } X_3: X_{31} \in [0, 0.3]; X_{32} \in [0.3, 0.6]; X_{33} \in [0.6, 1]$$

现在给定一组自变量数值 X : ($X_1 = 0.0871$; $X_2 = 0.4930$; $X_3 = 0.6598$) 让我们来判断图像中的天气是雾霾还是沙尘天气？

显然，我们的任务是判断图像中的天气情况，确定所选图像天气情况 W 取值 (W_N 或 W_Q)。为此，我们拿出图像中天气的一类情况进行考虑(其他类别同此分析)，构建朴素贝叶斯网络分类器。因变量 W 的类别是 W_N ，其他 3 个自变量 X_1 、 X_2 、 X_3 分别代表色相均值，饱和度均值，亮度均值。根据朴素贝叶斯假设子结点之间没有任何的依赖关系，即自变量之间相互独立。

$$w_n(x_{ij})_{\text{map}} = \arg \cdot \max P(x_{1j}|w_n)P(x_{2j}|w_n)P(x_{3j}|w_n)$$

其中 j 属于自变量的属性，根据给定的一组自变量数据和之前我们对自变量的属性分类，我们可以将上式改写为：

$$w_n(x_{ij})_{\text{map}} = \arg \cdot \max P(w_n)P(x_{11}|w_n)P(x_{22}|w_n)P(x_{33}|w_n)$$

为计算 W_N 的值，需要从表 1 所示的集合中估计出概率： $P(W_N)$ 、 $P(x_{11}|W_N)$ 、 $P(x_{22}|W_N)$ 、 $P(x_{33}|W_N)$ 、 $P(W_Q)$ 、 $P(x_{11}|W_Q)$ 、 $P(x_{22}|W_Q)$ 、 $P(x_{33}|W_Q)$ 等。

其中部分算法如下：

$$P(w_N) = (61+1)/(100+2) = 0.6078$$

$$P(w_Q) = (39+1)/(100+2) = 0.3922$$

$$P(x_{11}|w_N) = (0+1)/(61+2) = 0.0159$$

$$P(x_{22}|w_N) = (3+1)/(61+2) = 0.0635$$

$$P(x_{33}|w_N) = (53+1)/(61+2) = 0.8571$$

$$P(x_{11}|w_Q) = (38+1)/(39+2) = 0.9512$$

$$P(x_{22}|w_Q) = (19+1)/(39+2) = 0.4878$$

$$P(x_{33}|w_Q) = (33+1)/(39+2) = 0.8293$$

我们可以将处理后的所有条件概率汇总于表 2，表 3。

所以有：

$$w_N = \arg \cdot \max P(w_N)P(x_{11}|w_N)P(x_{22}|w_N)P(x_{33}|w_N) = 0.0526 \times 10^{-2}$$

$$w_Q = \arg \cdot \max P(w_Q)P(x_{11}|w_Q)P(x_{22}|w_Q)P(x_{33}|w_Q) = 0.1509$$

可见，朴素贝叶斯分类器将该图像的天气现象归为沙尘 W_Q 类。将上述概率进行归一化，可得到朴素贝叶斯分类器对该图像天气现象的识别概率：

$$P_N = \frac{w_N}{w_N + w_Q} = 0.0034, \quad P_Q = \frac{w_Q}{w_N + w_Q} = 0.9966$$

上述概率即为我们所修订后的后验概率，当然它们彼此之间没有可比性，因为历史资料告诉我们它

Table 2. The conditional probability of independent variables under the condition of haze W_N
表 2. 雾霾 W_N 条件下自变量各属性的条件概率

$P(x_{ij} W_N)$	X_{1j}	X_{2j}	X_{3j}
X_{11}	0.0159	0.9206	0.0159
X_{12}	0.9365	0.0635	0.1270
X_{13}	0.0476	0.0159	0.8571

Table 3. The conditional probability of independent variables under the condition of dust W_O
表 3. 沙尘 W_O 条件下自变量各属性的条件概率

$P(x_{ij} W_O)$	X_{1j}	X_{2j}	X_{3j}
X_{11}	0.9512	0.0488	0.0244
X_{12}	0.0244	0.4878	0.1463
X_{13}	0.0244	0.4634	0.8293

们各自出现的概率本来就不是平均的。所以我们再引入增长率 ζ ，它表示后验概率在其对应的先验概率的基础上的增量。此时我们在结合增长率之间的大小就可以对该天的天气现象做出相应的预报[10]-[15]。

如天气现象预报情况为：最有可能出现沙尘 W_O ，且出现的概率为 $P_O = 99.66\%$ 。而所选图像为沙尘天气，这与我们所得结果正好吻合，这也验证了朴素贝叶斯预报模型的适用性，为了进一步验证模型天气图像识别的准确性，我们用同样的方法选取多组自变量数值 X 进行检验，分别选取沙尘和雾霾的天气图像多张进行检验，准确率达到了 80% 以上。

5. 结论

本文提出了一种基于贝叶斯分类方法的天气图像自动识别方法。主要是对雾霾和沙尘两种天气图像进行识别分类，通过数字图像处理提取图像的色相、饱和度和亮度等特征，采用贝叶斯分类，建立贝叶斯分类模型，通过仿真验证了该模型的有效性，有比较高的准确性，准确率达到了 80% 以上。本方法可以适用于不同天气现象的识别，在天气现象识别方面可以取代人工观测，节省大量的人力物力，且本方法简单可靠、经济适用、精度较高，可以应用在船舶、机场等天气识别的场合。

基金项目

本工作得到国家自然科学基金项目(61002052, 61471412, 61250006)支持。

参考文献 (References)

- [1] 李骞, 范茵, 张璟, 李宝强 (2011) 基于室外图像的天气现象识别方法. *计算机应用*, **6**, 1624-1627.
- [2] 陈爱霞, 杜友福, 陈中举 (2015) 图像识别的应用现状和研究方向. *中国科技信息*, **16**, 1-2.
- [3] 徐峰 (2015) 基于数据库的叶片识别系统. *产业与科技论坛*, **12**, 42-43.
- [4] 张洁 (2014) 基于图像识别的农作物害虫诊断技术研究与应用. 博士论文, 中国科学技术大学, 合肥.
- [5] 蒋良孝 (2009) 朴素贝叶斯分类器及其改进算法研究. 博士论文, 中国地质大学, 武汉.
- [6] 董立岩, 苑森森, 刘光远, 贾书洪 (2007) 基于贝叶斯分类器的图像分类. *吉林大学学报(理学版)*, **2**, 249-253.
- [7] 陈法敬, 矫梅燕, 等 (2011) 亚高斯贝叶斯预报处理器及其初步试验. *气象学报*, **5**, 872-882.
- [8] Xu, G.L., Wang, X.T. and Xu, X.G. (2011) Improved bi-dimensional empirical mode decomposition based on 2d assisted signals analysis and application. *IET Image Processing*, **5**, 205-221. <http://dx.doi.org/10.1049/iet-ipr.2009.0158>
- [9] 彭丽玲 (2013) 基于图像灰度的天气变化机理及应用初步研究. 硕士论文, 昆明理工大学, 昆明.

- [10] Bayes, T.R. (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418. <http://dx.doi.org/10.1098/rstl.1763.0053>
- [11] Pearl, J. (1988) Probabilistic reasoning intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Mateo, 235-238.
- [12] Gonzalez, R.C. and Woods, R.E. (2002) Digital image processing. 2nd Edition, Prentice Hall, Upper Saddle River.
- [13] Katsura, H., Miura, J., Hild, M. and Shirai, Y. (2003) A view based outdoor navigation using object recognition robust to changes of weather and seasons. *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, October 2003, 2974-2979. <http://dx.doi.org/10.1109/IROS.2003.1249323>
- [14] Roser, M. and Moosmann, F. (2008) Classification of weather situations on single color images. 2008 *IEEE Intelligent Vehicles Symposium*, Eindhoven, 4-6 June 2008, 798-803. <http://dx.doi.org/10.1109/IVS.2008.4621205>
- [15] Yan, X., Luo, Y. and Zheng, X. (2009) Weather recognition based on images captured by vision system in vehicle. *Lecture Notes in Computer Science*, **5553**, 390-398. http://dx.doi.org/10.1007/978-3-642-01513-7_42