

文章编号: 1671-8860(2006)02-0108-04

文献标志码: A

基于 Naive Bayes Classifiers 的航空影像纹理分类

虞 欣¹ 郑肇葆¹ 汤 凌¹ 叶志伟¹

(1 武汉大学遥感信息工程学院, 武汉市珞喻路 129 号, 430079)

摘要: 在 Naive Bayes Classifiers 模型中, 要求父节点下的子节点(特征变量)之间相对独立, 然而在现实世界中, 特征与特征之间是非独立的、相关的。提出一种预处理方法, 实验结果表明, 该方法明显地提高了分类精度。

关键词: 贝叶斯网络; 纹理分类; 航空影像; 特征提取

中图法分类号: P237.3

影像纹理一直是计算机视觉、视觉生理学、心理物理学和认知科学等多学科研究的热点之一。近几年来, 出现了许许多新的纹理分类方法, 如基于神经网络、支撑向量机(SVM)和蚁群行为仿真的分类方法等。与医学影像、工业零件影像等相比, 航空影像类别较多, 像元混淆度大, 易出现错分和漏分的情况, 分类的难度较高。

人工智能的发展为 Bayes 理论的进一步发展提供了更为广阔的空间, 尤其是在 Bayes 学习理论方面, 其内容涵盖了人工智能的大部分领域^[1]。近年来, 贝叶斯网络成为研究的热点, 在人工智能领域大致经历了简单贝叶斯网络分类器(naive Bayes classifiers)^[2]、树增强型贝叶斯网络分类器(tree augmented naive Bayes classifiers)^[3]、一般贝叶斯网络分类器(general Bayesian network classifiers)和多个贝叶斯网络分类器(multinet Bayesian network classifiers)4个阶段^[4]。

在 Naive Bayes Classifiers 模型中, 要求父节点下的子节点之间相对独立, 然而在现实世界中并非如此。鉴于此, 本文根据特征间的相关性, 对特征进行一次合理的优化组合。该方法一方面降低了特征空间的维数, 避免了在高维空间计算的复杂性和不稳定性; 另一方面, 由于满足了 Naive Bayes Classifiers 的要求, 从而提高了分类的精度。

收稿日期: 2005-10-12。

项目来源: 国家自然科学基金资助项目(40271094)。

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

1 基于贝叶斯网络的分类法

贝叶斯网络结合了图形表示方法和概率知识, 是用来表示变量间连接概率的有向无环图, 它提供了一种自然的表示因果关系的方法(如图 1)。在这个网络中, 用节点 X_i ($i = 1, \dots, 6$) 表示变量, 有向边表示变量间的依赖关系, 即表示变量间的因果关系。从节点 X_i 到节点 X_j 的有向边的涵义是 X_i 对 X_j 有直接的因果影响。

贝叶斯网络 S 由一个二元组 $\langle G, P \rangle$ ^[4] 组成。其中, G 是有向无环图(directed acyclic graph, DAG); P 是局部概率分布的集合, $P = \{P(X|P_a(X)), X \in G\}$, $P_a(X)$ 代表节点 X 的父节点集合, 如 $P_a(X_5) = \{X_2, X_3\}$ 表示 X_2, X_3 都是 X_5 的父节点。

在上述贝叶斯网络定义的基础上, 如果假设在 G 中的具有同一子节点的父节点之间相互独立, 而且在同一层上的节点都相互独立, 便可大大地简化联合概率的计算。如果没有该假设, 图 1 中所有节点(如果为连续型随机变量)的联合概率为:

$$P(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(X_1, X_2, \dots, X_n) dX_1 dX_2 \dots dX_n \quad (1)$$

其中, $f(X_1, X_2, \dots, X_n)$ 为 n 维随机变量 X_1, X_2, \dots, X_n 的联合概率密度函数。在该假设下, 上述

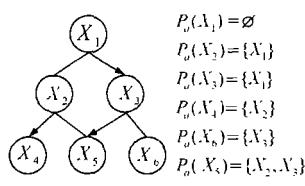


图1 贝叶斯网络的一个图例

Fig. 1 An Example of Bayesian Network

的联合概率有如下的简洁形式:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | P_a(X_i)) \quad (2)$$

图2表示在分类中的应用, 变量 A_1, \dots, A_n 为从某一待分类单元中提取的特征, n 为特征的个数, C 为类别变量。根据上述定义, 从图2中可以得到 $P_a(C) = 1$ 和 $P_a(A_i) = C, 1 \leq i \leq n$ 。如果假设同一层上的节点(A_1, A_2, \dots, A_n)相互独立, 则根据式(2)可得:

$$P(A_1, \dots, A_n, C) = P(C) \prod_{i=1}^n P(A_i | C) \quad (3)$$

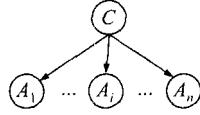


图2 在分类中的应用图例

Fig. 2 Example of Classification

依据条件概率的定义, 有:

$$P(C | A_1, \dots, A_n) = P(C) \prod_{i=1}^n \frac{P(A_i | C)}{P(A_i)} \quad (4)$$

其中, $P(C)$ 与 $P(A_i)$ 分别表示 C 类与特征 A_i 出现的先验概率, 而条件概率 $P(C | A_1, \dots, A_n)$ 表示在特征 A_1, A_2, \dots, A_n 出现的条件下, C 类出现的概率, 也称为 C 类的后验概率。同样, $P(A_i | C)$ 表示在 C 类条件下特征 A_i 出现的概率, $P(A_1, \dots, A_n, C)$ 表示 A_1, A_2, \dots, A_n 和 C 同时出现的联合概率。式(4)提供了由 C 类的先验概率和样本信息计算 C 类后验概率的简单计算方法^[2]。从式(4)中可以发现, 如果估计出 $P(C)$ 和 $P(A_i | C)$, 便可得到 $P(C | A_1, \dots, A_n)$ 。

2 Naive Bayes Classifiers 分类原理

本文采用灰度共生矩阵的5种统计性特征和小波分解的8个结构性特征。角二阶矩、对比度、相关、熵和逆差矩这5个特征依次记为 f_1, f_2, f_3, f_4 和 f_5 , 组成一个特征向量 F_s :

$$F_s = (f_1, f_2, f_3, f_4, f_5)^T \quad (5)$$

对输入图像进行小波(Symlets 小波)分解

分别提取一尺度的近似分量 LL 、水平细节分量 LH 、垂直细节分量 HL 和对角线细节分量 HH , 它们各自的均值和方差分别记为 $\mu_{LL}, \sigma_{LL}, \mu_{LH}, \sigma_{LH}, \mu_{HL}, \sigma_{HL}, \mu_{HH}, \sigma_{HH}$, 将这些特征值组成另一个特征向量 F_m :

$$F_m = (\mu_{LL}, \sigma_{LL}, \mu_{LH}, \sigma_{LH}, \mu_{HL}, \sigma_{HL}, \mu_{HH}, \sigma_{HH})^T \quad (6)$$

利用上述两个特征向量 F_s 和 F_m , 按 Naive Bayes Classifiers 原理分类的思路可以结合图3来说明。图3中, F 表示待分类单元的组合特征向量; F_s 和 F_m 为按式(5)、式(6)得到的特征向量; D_{s_i} 和 D_{m_i} (i 为类别数) 分别对应于 F 中的 F_s 和 F_m (即 $F = [F_s, F_m]$), 描述类别 C_i 的“代表性”特征向量。分类的目的就是计算后验概率 $P(C_i | F_s, F_m)$, 确定一个后验概率最大者, 相应的类别记为 C_i^* 。

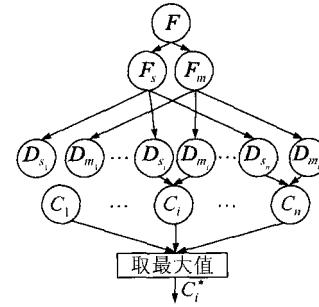


图3 Naive Bayes Classifiers 在纹理分类中的应用

Fig. 3 Naive Bayes Classifiers Applied in Texture Classification

后验概率为:

$$P(C_i | F_s, F_m) = P(F_s | D_{s_i}) P(D_{s_i}) + P(F_m | D_{m_i}) P(D_{m_i})^{[2]} \quad (7)$$

其中,

$$P(F_s | D_{s_i}) \propto \frac{1}{(2\pi)^{d/2} (|\Sigma_{s_i}|)^{1/2}} \cdot \exp\left[-\frac{1}{2} (F_s - \mu_{s_i})^T \Sigma_{s_i}^{-1} (F_s - \mu_{s_i})\right] \quad (8)$$

$$P(F_m | D_{m_i}) \propto \frac{1}{(2\pi)^{d/2} (|\Sigma_{m_i}|)^{1/2}} \cdot \exp\left[-\frac{1}{2} (F_m - \mu_{m_i})^T \Sigma_{m_i}^{-1} (F_m - \mu_{m_i})\right] \quad (9)$$

式中, μ_{s_i} 与 μ_{m_i} 分别为从训练样本提取统计性纹理特征的均值向量和样本方差矩阵; Σ_{s_i} 与 Σ_{m_i} 分别为结构性纹理特征的样本均值向量和样本方差矩阵; d 为特征向量的维数。

从式(8)和式(9)中可以看到, 由于 $F = [F_s, F_m]$, 所以特征向量的维数从 F 的维数降到 F_s 或 F_m 的维数。更为重要的是, 如果 F_s 或 F_m 内部特

特征间相关性较强, 则在计算 Σ_i 与 Σ_{m_i} 的行列式和逆时很不稳定, 不利于样本信息的充分提取, 所以有必要对特征进行相关性的分析, 进而对 F_s 和 F_m 进行重新分配和组合。

原算法中仅仅是根据纹理的属性(统计性和结构性)把 F 分为 F_s 和 F_m , 并未考虑到 F_s (或 F_m)内部特征间的相关性, 这往往影响求解的质量和分类的效果。考虑到这种情况, 并通过相关性分析打破统计性特征和结构性特征的界限, 对 F_s 和 F_m 进行重新分配, 取得合理的优化组合, 这便是对特征进行预处理的思路。为了书写的方便, 重新组合后的两个特征向量仍然采用 F_s 和 F_m 来描述, 显然它们失去了原先的“物理意义”。

假定 X 为 m 个训练样本提取 n 个特征的特征矩阵:

$$X = (X_1, \dots, X_n) = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix} \quad (10)$$

式中, x_{ij} 表示第 i 个训练样本的第 j 个特征的特征值; X_k 表示 m 个训练样本的第 j 个特征的特征值组成的列向量。在 n 个特征中既有统计性特征, 又有结构性特征。由 X 可得到特征间的相关系数矩阵 R 。

根据上述得到相关系数矩阵 R , 对大于某个经验值 Δ (Δ 一般取 $[0.3, 0.5]$, 过大或过小都会不同程度地影响分类的精度)的相关系数 r_{ij} 相对应的两个特征(i 和 j)进行分裂, 直到被分裂后的 F_s 或 F_m 中特征间的所有相关系数值都小于预先设定的经验值 Δ 。如果 r_{ij} 大于 Δ , 则把特征 i 与特征 j 分别置于 F_s 和 F_m 中, 不能同时把它们置于 F_s 或 F_m 中。从而可以保证重新被分裂的 F_s 或 F_m 内的特征与特征之间几乎两两相互独立——这便是对 F 进行分裂的原则。如此便使得矩阵 F_s 和 F_m 保持良好的状态, 同时也减少了由于特征间的相关性影响到最后的分类精度。这实际上是根据特征间的相关性对特征进行一次合理的优化组合。将上述分类方法归纳如下。

1) 从所有的样本中选取一定数量的样本作为训练样本, 在本文的实验中首先随机地(这是因为每类图像的编号是随机的)选取每类中前 6 幅图像作为训练样本;

2) 根据式(5)和式(6)提取训练样本中每一类的统计性特征 D_{s_i} 和结构性特征 D_{m_i} ;

3) 根据相关系数矩阵对特征进行合理的优化组合, 得到新的 D_{s_i} 、 D_{m_i} ;

4) 通过样本的先验信息和样本的特征信息

进行训练或学习, 即根据式(8)和式(9)可得到 $P(F_s | D_{s_i})$ 和 $P(F_m | D_{m_i})$ 中的 μ_{s_i} 、 μ_{m_i} 、 Σ_{s_i} 、 Σ_{m_i} ;

5) 根据式(5)和式(6)提取测试样本(在实验中包括训练样本)或 F_s 和 F_m ;

6) 根据训练阶段得到的分裂方法对 F_s 和 F_m 进行相对应的分裂, 得到新的 F_s 和 F_m ;

7) 利用式(7)计算出在 F_s 和 F_m 条件下 C 类出现的后验概率 $P(C_i | F_s, F_m)$;

8) 以后验概率最大原则对待分类单元进行分类或判别, 即 $C_i^* = \max_i \{P(C_i | F_s, F_m)\}$;

9) 计算总的分类精度, 如果分类精度未达到用户的需求, 则在原有的基础上逐步增加训练样本的数量, 重复 1)~8), 直到分类精度达到要求为止。当然, 也可借助于其他方法确定哪些样本作为训练样本。

3 实验结果与分析

本文实验数据是从航空影像中提取 141 幅 100×100 像素大小的图像, 根据野外调绘的结果, 将它们分成灌木、居民地、水田、河流和山地。实验结果见图 4 和表 1。

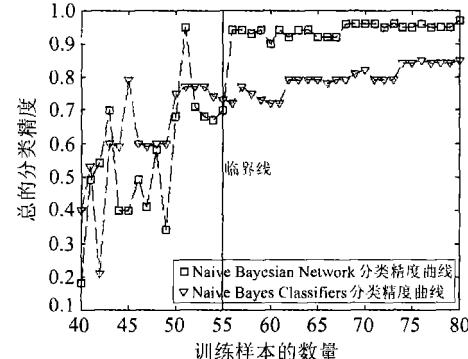


图 4 总的分类精度曲线图

Fig. 4 Curve of Total Classification Accuracy

表 1 分类结果统计表

Tab. 1 Statistics of Classification

N	I_NBC		NBC	
	A	M	A	M
56	0.94	9	0.72	39
57	0.94	9	0.77	33
58	0.93	10	0.75	35
59	0.94	9	0.73	38
60	0.90	14	0.72	39
61	0.94	9	0.72	39
62	0.92	11	0.79	30
63	0.94	9	0.79	29
64	0.94	8	0.79	29

1) 在图4中, 当训练样本多于55个时, 基于I_NBC的分类结果(总的分类精度)曲线明显地高于基于NBC的分类结果曲线。而且前者的分类精度曲线比后者的分类精度曲线更为平缓, 波动更小, 即当训练样本多于55个时, 前者的分类结果比后者更稳定。从另一个角度来看, 如果两种方法要达到同样的分类精度, 后者比前者需要更多的训练样本, 以便从那些“额外”增加的训练样本中提取“新”的样本信息来弥补由于特征间的相关性而损失的样本信息, 从而达到与前者同样的分类精度。此外, 前者只是增加特征相关性分析的时间。

2) 在表1(表1只列举了训练样本数在56与64之间两种分类方法的分类结果的比较)中, NBC代表Naive Bayes Classifiers, I_NBC代表经过预处理的NBC, N 表示训练样本的数量, A 代表总的分类精度, M 代表错分的样本数。从表1中可以发现, 当训练样本数在56与64之间时, I_NBC分类精度只是在[0.90, 0.96]这个很小的范围内变化, 也即当训练样本多于55个后, 过多增加训练样本的数量并不会给分类精度带来很明显的改善, 只会增加计算量和样本的费用。然而, 当训练样本数少

于40时, 分类精度波动很大, 况且分类精度也偏低, 其中最高的分类精度也只能达到0.54。因此, 选择多少样本, 以及选择哪些样本作为训练样本有待于进一步研究和探讨。

参 考 文 献

- [1] 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2002
- [2] Huang Yong, Chen Kapluck, Zhang Zhihua. Texture Classification by Multi-model Feature Integration Using Bayesian Networks[J]. Pattern Recognition Letters, 2003(24): 393-401
- [3] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers[J]. Mach. Learn., 1997(29): 131-163
- [4] Bart B, Geert V. Bayesian Network Classifiers for Identifying the Slope of the Customer Lifecycle of Long-life Customers[J]. European Journal of Operational Research, 2004(156): 508-523

第一作者简介: 虞欣, 博士生, 现主要从事图像解译和贝叶斯统计研究。

E-mail: china_yuxin@yahoo.com.cn

Aerial Image Texture Classification Based on Naive Bayes Classifiers

YU Xin¹ ZHENG Zhaobao¹ TANG Ling¹ YE Zhiwei¹

(1 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: In this paper, a new method, aerial images texture classification based on naive Bayes classifiers, is proposed. It is concluded that the new method performs better in overall classification precision than naive Bayes classifiers on the basis of the experiment results.

Key words: Bayesian network; texture classification; aerial image; features extraction

About the first author: YU Xin, Ph.D candidate, majors in image interpretation and Bayesian statistics.

E-mail: china_yuxin@yahoo.com.cn