Chapter I

A Novel Discriminative Naive Bayesian Network for Classification

Kaizhu Huang, Fujitsu Research and Development Centre Co. Ltd., China
Zenglin Xu, Chinese University of Hong Kong Shatin, Hong Kong
Irwin King, Chinese University of Hong Kong Shatin, Hong Kong
Michael R. Lyu, Chinese University of Hong Kong Shatin, Hong Kong
Zhangbing Zhou, Bell-Labs, Lucent Technologies, China

Abstract

Naive Bayesian network (NB) is a simple yet powerful Bayesian network. Even with a strong independency assumption among the features, it demonstrates competitive performance against other state-of-the-art classifiers, such as support vector machines (SVM). In this chapter, we propose a novel discriminative training approach originated from SVM for deriving the parameters of NB. This new model, called discriminative naive Bayesian network (DNB), combines both merits of discriminative methods (e.g., SVM) and Bayesian networks. We provide theoretic justifications, outline the algorithm, and perform a series of experiments on benchmark real-world datasets to demonstrate our model’s advantages. Its performance outperforms NB in classification tasks and outperforms SVM in handling missing information tasks.
Introduction

Bayesian network classifiers, a school of generative classifiers, have shown their advantages in many classification tasks, even though their overall performance is not as good as discriminative classifiers, such as support vector machines (Huang, Yang, King, & Lyu, 2004; Vapnik, 1999). The naive Bayesian network (NB) classifier is a simple yet effective Bayesian network classifier (Duda & Hart, 1973; Langley, Iba, & Thompson, 1992).

NB assumes a conditional independency among the variables or attributes. When used for classification, NB predicts a new data point as the class with the highest posterior probability. This is shown in equation (1), where $A_1, A_2, \ldots, A_n$ represents the attribute or variable, and $C_1, C_2, \ldots, C_k$ denotes the class variable. In equation (2), this posterior classification rule can be transformed into a joint probability classification rule, because $P(A_1, A_2, \ldots, A_n)$ for a given data point is a constant with respect to $C$. Finally, by incorporating the independency assumption, that is, $P(A_i, A_j | C) = P(A_i | C) P(A_j | C)$, for $1 \leq i \neq j \leq n$, the classification rule is changed in a decomposable form as equation (3).

$$
c = \arg \max_{C_i} P(C_i | A_1, A_2, \ldots, A_n)
= \arg \max_{C_i} \frac{P(C_i) P(A_1, A_2, \ldots, A_n | C_i)}{P(A_1, A_2, \ldots, A_n)}
= \arg \max_{C_i} P(C_i) \prod_{j=1}^{n} P(A_j | C_i)
$$

When used in real applications, NB first partitions the dataset into several subdatasets by the class label. Then, in each subdataset labelled by $C_i$, the maximum likelihood (ML) estimator $P(A_j = a_{jk} | C)$ can be given by the frequency $n_{ijk} / n_i$, $n_{ijk}$ is the number of the occurrences of the event $\{A_j = a_{jk}\}$ in subdataset $C_i$, $n_i$ is the number of the samples in subdataset $C_i$.

The above simple scheme achieves surprising success in many classification tasks (Duda & Hart, 1973; Friedman, Geiger, & Goldszmidt, 1997; Langley et al., 1992). Importantly, a great advantage of NB is its immediate ability to deal with the missing information problem. Assume the attributes set $\{A_1, A_2, \ldots, A_n\}$ be $A$. When the values of a subset of $A$, for example $T$, are unknown or missing, the marginalization inference can be obtained immediately as follows:

$$
c = \arg \max_{C_i} P(C_i) P(A - T | C_i)
= \arg \max_{C_i} P(C_i) P(A - T | C_i)
= \arg \max_{C_i} P(C_i) \prod_{j \in A - T} P(A_j | C_i)
$$

Copyright © 2007, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global. is prohibited.
An Improved Fuzzy C-Means Algorithm for Cell Formation Problems with Alternative Routes
www.igi-global.com/article/an-improved-fuzzy-c-means-algorithm-for-cell-formation-problems-with-alternative-routes/133124?camid=4v1a