Pattern Recognition 51 (2016) 358–377

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# SODE: Self-Adaptive One-Dependence Estimators for classification

Jia Wu [a,*], Shirui Pan [a], Xingquan Zhu [b], Peng Zhang [a], Chengqi Zhang [a]

[a] Quantum Computation & Intelligent Systems (QCIS) Centre, FEIT, University of Technology Sydney, NSW 2007, Australia
[b] Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

ARTICLE INFO

*Article history:*
Received 7 January 2015
Received in revised form
6 August 2015
Accepted 25 August 2015
Available online 5 September 2015

*Keywords:*
Attribute weighting
Self-adaptive
Naive Bayes
Classification
Artificial immune systems
Evolutionary machine learning

ABSTRACT

SuperParent-One-Dependence Estimators (SPODEs) represent a family of semi-naive Bayesian classifiers which relax the attribute independence assumption of Naive Bayes (NB) to allow each attribute to depend on a common single attribute (superparent). SPODEs can effectively handle data with attribute dependency but still inherent NB's key advantages such as computational efficiency and robustness for high dimensional data. In reality, determining an optimal superparent for SPODEs is difficult. One common approach is to use weighted combinations of multiple SPODEs, each having a different superparent with a properly assigned weight value (*i.e.*, a weight value is assigned to each attribute). In this paper, we propose a self-adaptive SPODEs, namely SODE, which uses immunity theory in artificial immune systems to automatically and self-adaptively select the weight for each single SPODE. SODE does not need to know the importance of individual SPODE nor the relevance among SPODEs, and can flexibly and efficiently search optimal weight values for each SPODE during the learning process. Extensive experiments and comparisons on 56 benchmark data sets, and validations on image and text classification, demonstrate that SODE outperforms state-of-the-art weighted SPODE algorithms and is suitable for a wide range of learning tasks. Results also confirm that SODE provides an appropriate balance between runtime efficiency and accuracy.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Naive Bayes (NB) [13] is a simple, efficient, and effective learning algorithm which uses a simplified Bayesian network, as shown in Fig. 1(a), with conditional attribute independence assumption for classification [18]. Despite of the strong independence assumption, NB has demonstrated very good classification accuracy, compared to many sophisticated learning methods [51]. Meanwhile, some efforts also exist to improve NB by relaxing its attribute interdependence but also retaining its simplicity and efficiency [26,47,48,53,54]. In this paper, we refer to this type of approaches as semi-naive Bayesian methods.

In order to relax the conditional independence assumption in naive Bayes and allow interdependency between attributes, semi-naive Bayesian techniques commonly employ simple wrapper heuristics to minimize learning errors on training data [54]. For example, One-Dependence Estimator (ODE), as an alternative to NB, allows each attribute to depend on at most one other attribute in addition to the class label, as shown in Fig. 1(b). Existing analysis and empirical studies [47] have shown that ODE can indeed outperform simple NB when the attribute independence assumption is violated.

SuperParent-One-Dependence Estimator (SPODE), as shown in Fig. 1(c), is a subcategory of ODE which allows all attributes to depend on one superparent (*i.e.*, one attribute) in addition to the class label [49]. The employment of a superparent allows SPODE to retain same training efficiency as NB but with a higher classification accuracy. Due to the fact that superparent plays a major role in SPODE but finding a global optimal superparent is a challenging task, many existing SPODE methods employ an ensemble based approach by using each single attribute as a superparent to build an SPODE and combining multiple SPODEs for prediction. For example, Averaged One-Dependence Estimators (AODE) [40] combines all SPODEs that satisfy a minimum support constraint and estimate class conditional probabilities using averaging strategy (*i.e.*, each attribute is treated equally). This approach has demonstrated good classification accuracy with very little extra computational cost. In reality, attributes play different roles in learning tasks. A natural way to extend AODE is to assign attributes different weight values, which is referred to as weighted SPODEs (WSPODE/WAODE) [22].

In order to discover proper weight values for weighted SPODEs, many useful methods have been proposed to evaluate the

* Corresponding author. Tel.: +61 416387666; fax: +61 2 9514 4535.
*E-mail addresses:* jia.wu@student.uts.edu.au (J. Wu), shirui.pan@student.uts.edu.au (S. Pan), xzhu3@fau.edu (X. Zhu), peng.zhang@uts.edu.au (P. Zhang), chengqi.zhang@uts.edu.au (C. Zhang).

http://dx.doi.org/10.1016/j.patcog.2015.08.023
0031-3203/© 2015 Elsevier Ltd. All rights reserved.

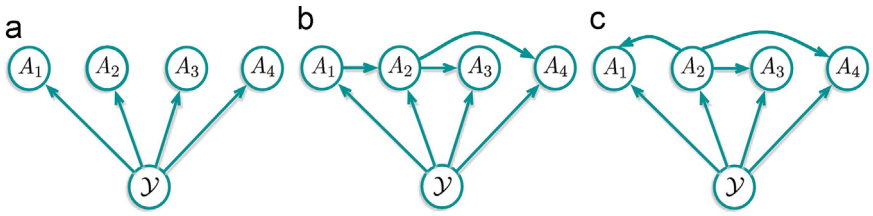


**Fig. 1.** A conceptual view of (a) Naive Bayes (NB), (b) One-Dependence Estimator (ODE), and (c) SuperParent-One-Dependence Estimator (SPODE). Each circle represents an attribute (*e.g.*, class label $\mathcal{Y}$ or attribute $A_i : 1 \le i \le 4$). An arrow points from a parent to a child, who only depends on its parents. NB assumes that attributes ($A_i$) are independent of each other given the class label $\mathcal{Y}$. ODE allows each attribute to depend on at most one other attribute (*i.e.*, parent) in addition to the class. By contrast, SPODE assumes that each attribute depends on a common attribute (*e.g.*, the superparent $A_2$).

importance of attributes. Examples include gain ratio [52], correlation-based algorithm [16], mutual information [24], and ReliefF attribute ranking algorithm [34]. Although existing attribute weighting SPODEs methods have achieved good performance to solve domain specific problems, all these methods rely on external criteria, such as gain ratio, to determine weight values of the attributes. In this case, attribute weighting and SPODE learning objective are separated without being considered simultaneously for maximum accuracy gain. To this end, we propose in this paper a new approach to automatically calculate optimal attribute weight values for SPODEs, by directly targeting SPODEs's objective function. To achieve the goal, we propose to assign proper weight values for weighted SPODEs classification based on immunity theory in artificial immune systems (AIS) [55]. Immune theory has been successfully used to self-adaptively calculate the weight for weighted naive Bayes in previous work [44]. In [45], an immune theory based self-adaptive probability estimation method has also been proposed to select terms and parameters for probability estimation. Therefore, it is appealing to the pattern recognition community to study an optimization framework with self-adaptively determined weight values for different learning tasks.

In this paper, we propose to use immune principle to design an automated searching strategy to find optimal attribute weight for each SPODE. The unique immune evolutionary computation processes, including initialization, clone, mutation, and selection, ensure that our method can adapt to the unique distributions of the underlying data. In contrast to conventional statistical probabilistic evaluation in SPODEs, the proposed immune based SPODEs (SODE) is a self-learning algorithm with immunological properties, such as memory property and clonal selection. To the best of our knowledge, this is the first work to introduce immune principle to SPODE based classification. The niche of SODE stem from the following three aspects:

(1) SODE is a data-driven self-adaptive method. It does not require explicit specification of functional or distributional form for the underlying learning models or tasks.
(2) SODE is a nonlinear model capable of modeling complex real-world relationships.
(3) SODE inherits the memory property of human immune systems and can recognize the same or similar antigen quickly at different times.

Our experiments and comparisons on 56 UCI benchmark data sets and validations on image and text classification demonstrate that SODE consistently outperforms other state-of-the-art weighted SPODEs algorithms in terms of classification accuracy and variance (*i.e.*, the standard deviation). The runtime comparisons further confirm that SODE provides an appropriate trades-off between learning efficiency and accuracy effectiveness.

The remainder of the paper is structured as follows. Preliminary concepts are addressed in Section 2. Section 3 presents an overview of attribute weighting approaches for SPODE classifiers, followed by a brief review of immune principle in artificial immune systems. Section 4 introduces the proposed algorithm, followed by experiments in Section 5. We conclude the paper in Section 6.

## 2. Preliminaries

In this section, we introduce important notations and definitions used in the paper.

A training set $\mathcal{D} = \{(\boldsymbol{x}_1, y_1)\cdots, (\boldsymbol{x}_N, y_N)\}$ has $N$ instances, each of which contains $n$ attribute values and a class label $y_i$. We use $\boldsymbol{x}_i = \{x_{i,1}, \cdots x_{i,j}, \cdots x_{i,n}\}$ to denote the $i$th instance in the data set $\mathcal{D}$, with $x_{i,j}$ denoting the $j$th attribute value and each instance is paired with a class label $y_i$. The class space $\mathcal{Y} = \{c_1, \ldots, c_k, \ldots, c_L\}$ denotes the set of labels each instance belonging to (*i.e.* $y_i \in \mathcal{Y} = \{c_1, \ldots, c_k, \ldots, c_L\}$) and $c_k$ is the $k$th label of the class space. The attribute space of the data is denoted by $\mathcal{A} = \{A_1, \ldots, A_j, \ldots, A_n\}$, where $A_j$ denotes the $j$th attribute. Each attribute can be a discrete random variable (with a number of discrete values) or a continuous random variable. In this paper, we only focus on categorical (or nominal) attributes, and for any attribute $A_j$, we use $a_j^\tau, \tau = 1, \ldots, |A_j|$ to denote the $\tau$th attribute value of $A_j$, and $|A_j|$ denotes the total number of distinct values of $A_j$. For each instance $\boldsymbol{x}_i$, its value satisfies $x_{i,j} \in A_j$.

*SPODE-based classifiers*: For an instance $(\boldsymbol{x}_i, y_i)$ in the training set $D$, its class label satisfies $y_i \in \mathcal{Y}$, whereas a test instance $\boldsymbol{x}_t$ only contains attribute values and its class label $y_t$ needs to be predicted by the classification model.

A Maxim A Posteriori (MAP) classifier aims to determine the class label of a test instance $\boldsymbol{x}_t$ by maximizing the posteriori probability as follows:

$$c(\boldsymbol{x}_t) = \underset{c_k \in \mathcal{Y}}{\operatorname{argmax}}\, P(c_k \mid \boldsymbol{x}_t) = \underset{c_k \in \mathcal{Y}}{\operatorname{argmax}}\, P(\boldsymbol{x}_t, c_k) \tag{1}$$

Because $P(c_k \mid \boldsymbol{x}_t) = P(c_k, \boldsymbol{x}_t)/P(\boldsymbol{x}_t)$ and $P(\boldsymbol{x}_t)$ is invariant across different class labels, one only needs to calculate $P(c_k, \boldsymbol{x}_t)$ to determine the final class label. In reality, when the number of training samples is limited, the estimation of joint distribution $P(c_k, \boldsymbol{x}_t)$ is usually unreliable. Therefore, approximating $P(c_k, \boldsymbol{x}_t)$ becomes the key challenge of deriving Bayesian learning models [23].

### 2.1. Naive Bayes

In reality, because joint probability $P(c_k, \boldsymbol{x}_t) = P(\boldsymbol{x}_t \mid c_k) \times P(c_k)$, a straightforward approach to simplify the joint probability estimation is to simply ignore dependency relationships between attributes and assume all attributes are conditionally independent, given the class label $c_k$. By doing so, the probability of observing the conjunction of all attributes is simplified as the product of the conditional probabilities of each individual attributes, which

Download English Version:

https://daneshyari.com/en/article/6940024

Download Persian Version:

https://daneshyari.com/article/6940024

Daneshyari.com