2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Attribute Weighting: How and When Does It Work for Bayesian Network Classification

Jia Wu, Zhihua Cai, Shirui Pan, Xingquan Zhu, and Chengqi Zhang

Abstract— **A Bayesian Network (BN) is a graphical model which can be used to represent conditional dependency between random variables, such as diseases and symptoms. A Bayesian Network Classifier (BNC) uses BN to characterize the relationships between attributes and the class labels, where a simplified approach is to employ a conditional independence assumption between attributes and the corresponding class labels, *i.e.,* the Naive Bayes (NB) classification model. One major approach to mitigate NB's primary weakness (the conditional independence assumption) is the attribute weighting, and this type of approach has been proved to be effective for NB with simple structure. However, for weighted BNCs involving complex structures, in which attribute weighting is embedded into the model, there is no existing study on whether the weighting will work for complex BNCs and how effective it will impact on the learning of a given task.**

In this paper, we first survey several complex structure models for BNCs, and then carry out experimental studies to investigate the effectiveness of the attribute weighting strategies for complex BNCs, with a focus on Hidden Naive Bayes (HNB) and Averaged One-Dependence Estimation (AODE). Our studies use classification accuracy (ACC), area under the ROC curve ranking (AUC), and conditional log likelihood (CLL), as the performance metrics. Experiments and comparisons on 36 benchmark data sets demonstrate that attribute weighting technologies just slightly outperforms unweighted complex BNCs with respect to the ACC and AUC, but significant improvement can be observed using CLL.

I. INTRODUCTION

BAYESIAN networks (BNs), consisting of a directed acyclic graph (DAG) and a set of local distributions [1], provide a means of expressing joint probability distributions over interrelated random variables. A node in the network corresponds to a variable, and the conditional probability table (CPT) associated to the node contains the probability of each state of the variable given every possible combination of states of its parents. The structure of the problem domain can be exploited by the explicit representation of probabilistic relations in the BN. In this way, incorporating domain knowledge into a BN model can be easily achieved. Besides, the intuitive graphical representation of the BN is very beneficial in decomposing a large and complex problem representation into several small and self-contained models.

In order to use BN for classification, the underlying Bayesian network classifier (BNC) from a given set of labeled training instances that are represented by a tuple of attribute variables should be constructed in order to predict the distribution of the class variable. Learning a BNC is an established research topic since the past decade. Generally, the Bayesian approach for classification is to assign the most probable target value to the test instance. Typically, one set of training instances with class labels are given, a classifier must be learned to predict the class distribution of an instance with unknown class label. Assume that all attributes satisfy the conditional attribute independence assumption, the probability of observing the conjunction is just the product of the probabilities for the individual attributes. This is the essential concept of naive Bayes, NB, a highly practical Bayesian network classification method. The naive Bayes classification is based on the so-called Bayesian theorem and is particularly suitable for high dimensional data. In naive Bayes, each attribute node has the class node as its only parent, and it does not have any other parent from other attribute nodes, as shown in Figure 1.

In reality, the strong conditional attribute independence assumption made by NB may reduce its classification performance when the condition is violated in a learning task. To alleviate the attribute independence assumption of NB and retain NB's simplicity and efficiency, researchers have proposed many effective methods to further improve its performance. The following three methods have demonstrated good accuracy. Selective naive Bayes (SBC) [2] demonstrates a significant improvement by using the selected subset of variables, which optimizes the classification accuracy. Tree augmented naive Bayes (TAN) [3] naturally extends the naive Bayes classifier, as shown in Figure 2. Naive Bayes/Decision-Tree Hybrid (NBTree) [4] combines a decision tree with NB. Overall, these techniques have high computational overheads

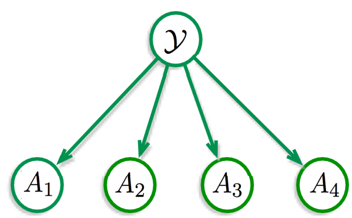

Fig. 1. The structure of Naive Bayes (NB).

Jia Wu is with the Centre for Quantum Computation & Intelligent Systems, FEIT, University of Technology Sydney, Australia, and is also affiliated with the Department of Computer Science, China University of Geosciences Wuhan, China (jia.wu@student.uts.edu.au).

Zhihua Cai is with the Department of Computer Science, China University of Geosciences Wuhan, China (zhcai@cug.edu.cn).

Xingquan Zhu is with Dept. of Computer and Electrical Engineering & Computer Science, Florida Atlantic University, USA (xzhu3@fau.edu).

Shirui Pan and Chengqi Zhang are with the Centre for Quantum Computation & Intelligent Systems, FEIT, University of Technology Sydney, Australia ({shirui.pan@student.,chengqi.zhang@}uts.edu.au).

The work was supported by the National Scholarship for Building High Level Universities, China Scholarship Council (No. 201206410056).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

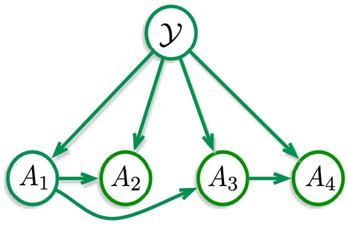

Fig. 2. The structure of Tree Augmented Naive Bayes (TAN).

during training or classification phases: (a) TAN and NBTree have high computational complexity at training time; and (b) SBC has high computational complexity at classification time.

The high computational overheads of the above methods motivates the development of averaged one-dependence estimators AODE [5]. As shown in Figure 3, an one-dependence classifier is firstly trained for each attribute, in which the attribute is set to be the parent of all other attributes. Then, AODE directly averages the aggregation consisting of many special tree augmented naive Bayes. In addition to having good classification performance, AODE retains the simplicity and direct theoretical foundation of NB without incurring high computational costs. Su & Zhang [6] explored and represented variable independence in learning conditional probability tables (CPTs), and then proposed a full Byaesian network classifier (FBC). Recently, Zhang & Jiang [7] proposed a high-performance BNC, named hidden naive Bayes (HNB). In HNB, the authors used the conditional mutual information as the weight of the hidden parent attribute, as shown in Figure 4.

Another major way to help mitigate its primary weakness (*i.e.* attributes independence assumption) is to assign larger weights to important attributes in classification. This is mainly because attributes do not play the same role in different learning tasks, and some attributes are more important than others for a specific learning task. For NB with a simple structure, a natural way to extend NB is to assign attributes different weight values to relax the conditional independence assumption, called weighting naive Bayes, simply WNB. For example, Jiang [8] proposed to use weight averaged one-dependence estimators WAODE to improve AODE model.

In order to learn proper weight values for BNC, researchers have proposed many methods to evaluate the importance of attributes, including Gain Ratio [9], CFS (Correlation-based Feature Selection) attribute selection algorithm [10], Mutual Information [7], [8], ReliefF attribute ranking algorithm [11] and so on. Besides, Hall [12] proposed a novel attribute weighting methods based on the degree to which they depended on the values of other attributes.

The above attribute weighting methods for NB have achieved good performance to solve domain specific problems. However, for complex BNC models, such as AODE and HNB, attribute weighting receives very little attention. Moreover, for the existing WAODE [8], besides the M-estimation the underlying attribute weighting method is only a part of the improvements, similar to HNB. In this case, there is no comprehensive understanding on how much impact the attribute weighting can bring to the classification results. In this paper, we analyze the performance of attribute weighted complex BNCs (WAODE and HNB) by using different attribute weighting methods. Experiments and comparisons, on 36 UCI benchmark data sets [13] demonstrate that attribute weighting technologies only slightly outperform unweighted complex BNCs on classification accuracy (ACC) and area under the ROC curve ranking (AUC) [14]. However, significant improvement can be observed using conditional log likelihood (CLL) [15].

II. Attribute Weighted BNCs

Given a training set $\mathcal{D} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ with N instances, each of which contains n attribute values and a class label. We use $\mathbf{x}_i = \{x_{i,1}, \cdots x_{i,j}, \cdots x_{i,n}, y_i\}$ to denote the ith instance in the data set $\mathcal{D}$, with $x_{i,j}$ denoting the jth attribute value and y_i denoting the class label of the instance. The class space $\mathcal{Y} = \{c_1, \cdots, c_k, \cdots, c_L\}$ denotes the set of labels that each instance belongs to and c_k denotes the kth label of the class space. For ease of understanding, we use (x_i, y_i) as a shorthand to represent an instance and its class label, and use x_i as a shorthand of $\mathbf{x}_i$. We also use A_j as a shorthand to represent the jth attribute. Each attribute can be a discrete random variable (with a number of discrete

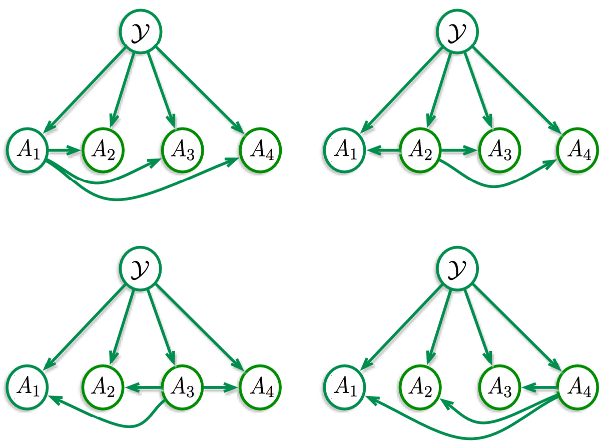

Fig. 3. The structure of Averaged One-Dependence Estimators (AODE).

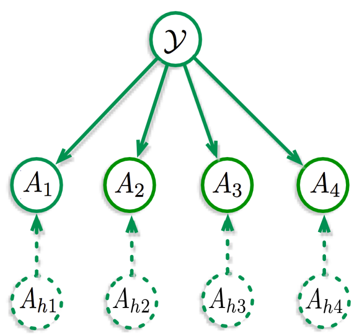

Fig. 4. The structure of Hidden Naive Bayes (HNB).

TABLE I

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED WAODE VS. AODE: ACC AND STANDARD DEVIATION.

Data Sets	AODE	CFS-WAODE	GR-WAODE	MI-WAODE	Ref-WAODE	DT-WAODE
anneal	96.83±1.66	96.92±1.59	96.90±1.56	97.38±1.58	97.54±1.51	97.15±1.59
anneal.ORIG	89.01±3.10	89.40±3.09	89.59±3.07	89.70±2.78	89.86±2.86	89.59±2.95
audiology	71.66±6.42	71.70±6.48	71.66±6.42	71.61±6.67	71.57±6.59	71.65±6.54
autos	74.60±10.10	75.18±10.26	75.08±10.03	75.43±9.96	75.32±10.00	76.20±10.22
balance-scale	89.78±1.88	88.32±2.34 ○	89.65±2.00	89.65±2.00	89.15±2.21	89.71±1.97
breast-cancer	72.73±7.01	72.77±6.88	71.80±7.33	72.25±7.11	72.53±7.13	72.67±7.11
breast-w	96.85±1.90	96.68±1.94	96.82±1.91	96.82±1.91	96.81±1.91	96.67±2.05
colic	80.93±6.16	81.01±6.23	81.36±6.03	81.50±6.09	81.72±5.90	81.23±6.33
colic.ORIG	75.38±6.41	75.60±6.51	75.71±6.21	76.26±6.35	75.82±6.46	75.85±6.21
credit-a	85.86±3.72	85.96±3.73	85.97±3.58	85.90±3.57	85.90±3.57	85.90±3.76
credit-g	76.45±3.88	76.33±3.80	76.52±3.62	76.30±3.63	76.20±3.75	76.45±3.77
diabetes	76.57±4.53	76.42±4.58	76.33±4.64	76.20±4.64	76.05±4.70	76.54±4.52
glass	61.73±9.69	61.87±9.43	62.06±9.46	61.40±9.38	61.26±9.24	61.78±9.68
heart-c	82.84±7.03	82.97±6.91	83.04±6.93	83.04±6.83	83.10±6.91	83.00±7.00
heart-h	84.09±6.00	84.32±5.77	84.29±5.52	84.50±5.77	84.30±5.92	84.36±6.07
heart-statlog	83.63±5.32	83.63±5.67	83.52±5.80	83.93±5.84	83.48±5.90	83.59±5.89
hepatitis	85.21±9.36	84.76±9.65	85.09±9.63	84.06±10.05	85.15±9.49	85.02±9.51
hypothyroid	93.56±0.61	93.61±0.56	93.58±0.55	93.52±0.56	93.58±0.54	93.62±0.57
ionosphere	91.74±4.28	91.99±4.17	91.88±4.13	91.85±4.12	91.74±4.23	91.85±4.01
iris	94.00±5.88	94.40±5.50	95.07±5.16	95.00±5.14	95.40±4.80	94.67±5.53
kr-vs-kp	91.03±1.66	92.30±1.46 •	93.26±1.39 •	94.14±1.28 •	93.51±1.45 •	91.64±1.71 •
labor	94.57±9.72	94.93±9.13	94.93±9.13	94.03±10.18	94.17±9.90	94.17±9.48
letter	77.64±2.02	78.06±2.07 •	78.65±1.93 •	78.70±1.94 •	78.87±1.95 •	77.80±2.08
lymph	85.46±9.32	85.39±9.24	85.59±9.20	85.46±9.23	85.91±9.45	85.92±9.21
mushroom	99.94±0.19	99.92±0.21	99.88±0.24	99.87±0.25	99.90±0.23	99.68±0.42
primary-tumor	47.87±6.37	47.87±6.46	47.96±6.47	47.70±6.42	47.61±6.55	47.72±6.44
segment	92.92±1.40	92.89±1.42	92.98±1.44	93.13±1.49	93.22±1.45	92.83±1.43
sick	97.52±0.72	97.59±0.71	97.63±0.75	98.01±0.72 •	97.84±0.69 •	97.74±0.66 •
sonar	79.91±9.60	80.20±9.38	80.20±9.52	80.30±9.44	80.91±9.16	79.92±9.38
soybean	93.31±2.85	93.38±2.78	93.28±2.87	93.26±2.82	93.32±2.74	93.32±2.80
splice	96.12±1.00	96.18±0.98	96.13±0.99	96.11±1.01	96.20±0.98	96.11±1.01
vehicle	71.65±3.59	71.58±3.58	71.85±3.61	71.83±3.64	71.70±3.62	71.64±3.67
vote	94.52±3.19	94.52±3.19	94.46±3.17	94.46±3.17	94.11±3.35	94.25±3.28
vowel	89.64±3.06	89.56±3.09	89.04±3.21	89.09±3.21	89.73±3.15	89.56±3.13
waveform-5000	84.84±3.07	84.78±3.24	85.19±3.11	85.18±3.14	85.04±3.15	85.00±3.08
zoo	94.66±6.38	93.76±6.43	94.66±6.38	93.76±6.43	93.96±6.46	94.57±6.50
Mean	84.86±4.70	84.91±4.93	85.04±4.81	85.04±4.98	85.07±4.96	84.98±4.71
w/t/l	-	2/33/1	2/34/0	3/33/0	3/33/0	2/34/0

•, ○: Statistically significant upgradation and degradation, respectively.

values) or a continuous random variable. For each instance $\mathbf{x}_i$, its value satisfies $x_{i,j} \in A_j$. For an instance (x_i, y_i) in the training set $\mathcal{D}$, its class label satisfies $y_i \in \mathcal{Y}$, whereas a test instance x_t only contains attribute values and its class label y_t needs to be predicted by the BNC models.

A. WNB: Attribute Weighted NB

Because attributes in NB may not play equally important roles in different learning tasks, a natural way to extend NB is to assign attributes different weight values to relax the conditional independence assumption. This is the main idea of the weighted naive Bayes (WNB) methods, which can be mathematically formulated as follows.

$$c(x_t) = \arg\max_{c_k \in \mathcal{Y}} P(c_k) \prod_{j=1}^{n} P(x_{t,j} \,|c_k)^{w_j} \tag{1}$$

In Eq. (1), $P(c_k)$ denotes the probability of class c_k in the whole training set. $P(x_{t,j} \,|c_k)$ denotes the joint distribution of $x_{t,j}$ conditioned by the the given class c_k, where w_j denotes the weight of jth attribute. In [16], an empirical analysis was carried out on WNB and the underlying results show that attribute weighting can significantly improve the performance of NB. Therefore, we do not include the experimental analysis for WNB in this paper.

B. WAODE: Attribute Weighted AODE

The main object of attribute weighting for averaged one-dependence estimators (WAODE), proposed by Jiang [8], is to improve the ACC performance of AODE. In the original design, a tree augmented naive Bayes was built for each attribute, in which the attribute was set as the root attribute. Each tree augmented naive Bayes was assigned a weight value determined by the mutual information between the attribute and the class variable, which could be formulated as

$$c(x_t) = \arg\max_{c_k \in \mathcal{Y}} \frac{\sum\limits_{j=1}^{n} w_j P(x_{t,j}, c_k) \prod\limits_{i=1}^{n} P(x_{t,i}|x_{t,j}, c_k)}{\sum_{j=1}^{n} w_j} \tag{2}$$
$$s.t. \qquad F(x_{t,j}) \geq m$$

where $F(x_{t,j})$ denotes the number of training instances having attribute-value $x_{t,j}$ and is used to enforce the limit m that they place on the support needed in order to accept a conditional probability estimation. n is the number of attributes. In order to avoid unreliable probability estimations, WAODE excludes models where the frequency of the value for classified object of the parent attribute in the training data is fewer than the limit (m=30).

C. HNB: Hidden Naive Bayes

Hidden Naive Bayes (HNB) is an extension of the naive Bayes classifier. It employs an approximation of the joint

TABLE II

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED WAODE VS. AODE: AUC AND STANDARD DEVIATION.

Data Sets	AODE	CFS-WAODE	GR-WAODE	MI-WAODE	Ref-WAODE	DT-WAODE
anneal	98.93±1.50	98.96±1.47	98.95±1.48	98.97±1.47	98.97±1.49	98.98±1.47
anneal.ORIG	97.29±3.93	97.38±3.76	97.43±3.69	97.56±3.30	97.65±3.08	97.47±3.50
audiology	83.81±1.48	83.82±1.48	83.82±1.48	83.84±1.48	83.86±1.49	83.85±1.49
autos	94.67±2.54	94.76±2.51	94.72±2.52	94.83±2.51	94.83±2.48	94.88±2.53
balance-scale	79.93±3.94	79.21±4.03 ○	79.81±3.95	79.81±3.95	79.70±4.00	79.82±3.95
breast-cancer	71.18±10.03	70.88±10.04	69.79±10.12	70.09±10.25	71.01±9.96	70.98±10.13
breast-w	99.28±0.73	99.27±0.76	99.28±0.73	99.29±0.73	99.29±0.73	99.27±0.75
colic	86.79±6.08	86.97±6.05	87.25±6.03	87.15±6.05	87.64±5.98	87.13±5.96
colic.ORIG	82.22±7.24	82.72±7.13	83.29±6.93	82.95±7.20	82.67±7.29	82.69±7.08
credit-a	92.35±2.92	92.39±2.93	92.47±2.93	92.39±2.94	92.47±2.91	92.40±2.94
credit-g	79.68±4.14	79.67±4.19	79.62±4.21	79.64±4.20	79.58±4.17	79.68±4.14
diabetes	82.96±4.83	83.02±4.76	83.08±4.71	83.05±4.68	82.89±4.71	83.03±4.82
glass	83.60±5.64	83.78±5.48	84.04±5.38	83.58±5.62	83.82±5.67	83.72±5.55
heart-c	84.11±0.57	84.09±0.58	84.08±0.58	84.08±0.58	84.08±0.58	84.09±0.58
heart-h	83.97±0.54	83.98±0.54	83.99±0.56	83.98±0.55	83.98±0.55	83.96±0.55
heart-statlog	91.28±4.70	91.24±4.68	91.18±4.71	91.14±4.64	91.19±4.70	91.22±4.68
hepatitis	88.53±10.56	88.42±10.74	88.57±10.45	87.80±10.84	88.22±10.75	88.44±10.42
hypothyroid	87.34±7.23	87.46±7.09	87.88±6.54	87.43±6.85	87.11±7.17	87.36±7.11
ionosphere	97.57±2.23	97.52±2.37	97.54±2.33	97.51±2.29	97.54±2.26	97.47±2.43
iris	99.16±1.42	99.15±1.47	99.17±1.43	99.17±1.43	99.17±1.43	99.13±1.44
kr-vs-kp	97.44±0.75	97.89±0.66 •	98.14±0.61 •	98.53±0.52 •	98.33±0.57 •	97.71±0.69 •
labor	98.54±4.90	98.38±5.65	98.50±5.54	98.50±5.54	98.08±5.93	98.08±6.19
letter	98.40±0.28	98.46±0.27 •	98.55±0.25 •	98.55±0.25 •	98.60±0.24 •	98.42±0.27 •
lymph	95.03±4.35	95.09±4.32	94.95±4.30	95.13±4.34	95.09±4.31	95.16±4.35
mushroom	100.00±0.01	100.00±0.01	100.00±0.01	100.00±0.01	100.00±0.01	99.99±0.02
primary-tumor	85.72±2.07	85.78±2.04	85.75±2.06	85.79±2.04	85.78±2.06	85.75±2.05
segment	99.42±0.22	99.42±0.22	99.44±0.21 •	99.46±0.21 •	99.47±0.21 •	99.42±0.21
sick	97.09±1.69	97.41±1.50 •	97.55±1.46 •	98.36±1.01 •	98.12±1.13 •	97.66±1.40 •
sonar	90.01±6.77	90.02±6.83	90.02±6.85	90.25±6.75	90.58±6.61	89.95±6.71
soybean	99.91±0.09	99.91±0.09	99.91±0.08	99.92±0.08	99.91±0.09	99.91±0.09
splice	99.56±0.25	99.56±0.25	99.55±0.26	99.55±0.26	99.56±0.25	99.56±0.25
vehicle	89.91±2.03	89.95±1.99	89.90±2.04	89.91±2.05	89.92±2.05	89.94±2.00
vote	98.67±1.24	98.67±1.24	98.77±1.21	98.77±1.21	98.55±1.30	98.59±1.31
vowel	99.40±0.36	99.39±0.36	99.36±0.36	99.36±0.35	99.40±0.35	99.40±0.36
waveform-5000	96.70±1.27	96.68±1.25	96.65±1.24	96.64±1.24	96.62±1.26	96.72±1.25
zoo	99.07±1.43	99.07±1.43	99.07±1.43	99.07±1.43	99.07±1.43	99.07±1.43
Mean	91.93±3.05	91.95±3.06	92.00±3.02	92.00±3.02	92.02±3.03	91.97±3.06
w/t/l	-	3/32/1	4/32/0	4/32/0	4/32/0	3/33/0

•, ○: Statistically significant upgradation and degradation, respectively.

distribution defined as follows.

$$c(x_t) = \arg\max_{c_k \in \mathcal{Y}} P(c_k) \prod_{j=1}^{n} P(x_{t,j}\,|A_{hj}, c_k) \tag{3}$$

where

$$P(x_{t,j}|A_{hj}, c_k) = \sum_{j=1, j\neq i}^{n} w_{i,j} P(x_{t,j}|x_{t,i}, c_k) \tag{4}$$

where $w_{i,j}$ is the conditional weight contributed by attribute A_i and A_j. In HNB, attribute dependencies are actually represented by hidden parents of attributes. It can be viewed in such a way that a hidden parent A_{hj} is created for each attribute A_j. One-dependence estimators $P(x_{t,j}|x_{t,i}, c_k)$ are used to define hidden parents. Presumably, HNB is an accurate model because it can represent the influences on each attribute from all other attributes and assign higher weights to important attributes.

D. Attribute Weighted Approach

The way to learn the attribute weights is the most important part for attribute weighted BNCs, which can be summarized as follows.

1) MI (Mutual Information): According to the probability and information theory, the mutual information of two random variables provides a quantified measure to evaluate the mutual dependence of two variables, which can be defined as

$$w_j = \sum_{a_j^{\tau} \in A_j, c_k \in \mathcal{Y}} P(a_j^{\tau}, c_k) log \frac{P(a_j^{\tau}, c_k)}{P(a_j^{\tau})P(c_k)} \tag{5}$$

Mutual information has been widely used for measuring the importance between attributes and the class variable in classification. For HNB, Jiang [7] proposed a conditional mutual information based method to calculate the $w_{i,j}$ defined as

$$w_{i,j} = \frac{I_{i,j}}{\sum_{j=1, j\neq i}^{n} I_{i,j}} \tag{6}$$

where $I_{i,j}$ is the conditional mutual information between $x_{t,i}$ and $x_{t,j}$ given the class c_k, which can be defined as

$$I_{i,j} = \sum_{x_{t,i}, x_{t,j}, c_k} P(x_{t,i}, x_{t,j}, c_k) log \frac{P(x_{t,i}, x_{t,j}|c_k)}{P(x_{t,i}|c_k)P(x_{t,j}|c_k)} \tag{7}$$

2) GR (Gain Ratio): Zhang & Sheng [17] argued that an attribute with a higher gain ratio [9] deserves a higher weight value in WNB. In their studies, they proposed a gain ratio weighted method that calculates the weight of an attribute from a data set, as shown in the following Eq. (8).

$$w_j = \frac{GainRatio(A_j) \times n}{\sum_{j=1}^{n} GainRatio(A_j)} \tag{8}$$

TABLE III

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED WAODE VS. AODE: CLL AND STANDARD DEVIATION.

Data Sets	AODE	CFS-WAODE	GR-WAODE	MI-WAODE	Ref-WAODE	DT-WAODE
anneal	-9.56±5.69	-9.43±5.65	-9.43±5.65	-9.13± 5.68	-9.05±5.70	-9.55±5.66
anneal.ORIG	-25.08±5.92	-24.72±5.91 •	-24.42±5.91 •	-24.48±5.95	-24.30±5.95 •	-24.61± 5.95
audiology	-286.15±102.49	-285.89±102.19	-285.59±102.58 •	-286.28±101.61	-286.32±101.63	-285.85±102.02
autos	-133.92±61.13	-132.07±61.60	-133.23±61.30	-130.85±61.48 •	-130.34±61.22 •	-129.30±60.88 •
balance-scale	-53.21±2.50	-53.86±2.57 ○	-53.36±2.53 ○	-53.36±2.53 ○	-53.36±2.58	-53.36±2.53 ○
breast-cancer	-58.49±11.77	-58.31±11.27	-58.49±10.55	-58.35±10.68	-59.43±12.24 ○	-58.43±11.57
breast-w	-10.63±6.92	-10.81±7.04	-10.71±6.98	-10.62±6.91	-10.66±6.91	-10.78±6.95
colic	-55.35±18.75	-55.12±18.71	-54.93±18.76	-54.41±18.62	-54.41±18.82	-54.97±18.75
colic.ORIG	-54.23±12.78	-53.65±12.82	-52.91±12.96	-54.11±13.28	-54.18±13.42	-53.75±12.81
credit-a	-38.40±10.06	-37.96±9.87 •	-37.43±9.81	-37.62±9.79	-37.73±9.97	-38.13±10.01
credit-g	-51.48±5.80	-51.48±5.86	-51.58±5.91	-51.83±6.01	-51.85±5.96	-51.48±5.81
diabetes	-49.86±8.00	-49.81±7.99	-49.88±8.04	-49.93±8.00	-50.11±7.98	-49.78±8.01
glass	-103.54±21.42	-103.13±21.96	-102.94±22.04	-103.87±21.81	-103.68±21.99	-103.25±21.60
heart-c	-42.72±18.57	-42.82±18.54	-42.63±18.48	-42.67±18.23	-43.12±18.78	-42.80±18.42
heart-h	-43.22±15.20	-42.37±14.59 •	-41.59±14.15	-41.43±13.68	-41.66±13.95	-42.69±14.46
heart-statlog	-44.25±17.03	-44.58±17.19	-44.59±17.28	-44.58±17.09	-45.05±17.37	-44.53±17.11
hepatitis	-42.96±27.80	-42.23±27.31	-41.64±26.94	-42.43±26.91	-42.21±27.10	-42.21±27.12
hypothyroid	-23.39±2.94	-23.15±2.89 •	-23.06±2.78 •	-23.16±2.92	-23.20±2.93	-23.19±2.90 •
ionosphere	-59.76±39.09	-59.63±39.30	-59.67±39.23	-60.07±39.42	-59.98±39.37	-60.06±40.11
iris	-16.35±9.37	-15.95±9.47	-15.66±9.35	-15.68± 9.36	-15.48±9.41	-16.10±9.73
kr-vs-kp	-24.45±1.86	-23.22±1.78 •	-22.00±1.78 •	-20.38±1.70 •	-20.96±1.80 •	-23.26±1.83 •
labor	-15.30±20.12	-16.03±21.09	-16.24±21.63	-16.49±21.35	-17.76±23.90	-18.65±25.16
letter	-79.52±7.19	-77.94±7.15 •	-75.67±7.07 •	-75.66±7.05 •	-74.65±7.08 •	-78.97±7.19 •
lymph	-38.24±23.93	-38.03±23.51	-37.76±23.31	-37.81±23.06	-37.79±23.25	-38.15±23.61
mushroom	-0.51±1.25	-0.54±1.19	-0.54±1.11	-0.57± 1.15	-0.57±1.30	-1.00±1.40 ○
primary-tumor	-188.96±22.30	-188.54±22.04	-188.56±22.12	-188.59±22.04	-188.95±22.17	-188.85±22.20
segment	-22.07±5.57	-22.06±5.53	-21.50±5.49 •	-21.06±5.43 •	-21.10±5.49 •	-22.14±5.52
sick	-8.55±2.37	-7.89±2.22 •	-7.69±2.20 •	-6.52±2.03 •	-6.86±2.08 •	-7.44±2.15 •
sonar	-67.79±35.58	-68.41±36.33	-68.68±36.74	-67.79±36.39	-68.41±37.33	-70.17±38.64
soybean	-22.90±9.62	-21.60±9.12 •	-22.73±9.61	-21.09±8.84 •	-21.03±8.84 •	-22.68±9.60 •
splice	-12.07±3.30	-11.95±3.31 •	-11.94±3.37	-11.94±3.36	-11.93±3.32 •	-12.01±3.33
vehicle	-79.40±11.54	-79.60±11.73	-78.96±11.58	-78.82±11.57	-79.17±11.59	-79.60±11.56
vote	-17.27±10.77	-16.73±10.20 •	-15.56±9.51 •	-15.51±9.49 •	-17.71±10.43	-17.72±10.70
vowel	-31.15±6.85	-31.33±6.89	-32.24±6.91	-32.15± 6.88	-30.82±6.80	-31.09±6.83
waveform-5000	-37.08±7.84	-36.50±7.69	-36.06±7.76 •	-36.06±7.76 •	-36.17±7.75	-36.56±7.78 •
zoo	-10.26±9.24	-9.68±8.58	-10.09±9.03	-9.60± 8.50	-9.84±8.76	-10.50±9.74
Mean	-51.6±16.18	-51.31±14.14	-51.11±16.12	-50.97±16.02	-51.11±16.25	-51.49±16.38
w/t/l	-	10/25/1	9/26/1	8/27/1	8/27/1	7/27/2

•, ○: Statistically significant upgradation and degradation, respectively.

3) CFS (Correlation-based Feature Selection): CFS based attribute weighting uses a correlation-based heuristic evaluation function as the attribute quality measure [10] to calculate weight of each attribute. The core of CFS algorithm is the heuristic process that evaluates the worth or "merit" of a subset of features. Hall [12] employed this method to evaluate the importance of attributes according to the heuristic "merit" value. The weight of attribute A_j can be defined as

$$w_j = \frac{1}{\sqrt{indexrank(A_j) + 1}} \tag{9}$$

where $indexrank(A_j)$ is the index of the ranked attributes according to the order of being added to the subset during a forward selection search in CFS.

4) Ref (Relief-F): Relief is a feature selection method based on attribute estimation [18]. Relief assigns a grade of relevance to each feature by examining the change of the feature values with respect to instances within the same class (*i.e.* the nearest hit) and instances between classes (*i.e.* the nearest miss). If a feature's values remain relatively stable for instances within the same class, the feature will receive a higher weight value. The original Relief only handles binary classification problems. Its extension, Relief-F, can be applied for multi-class classification [11].

5) DT (Decision Tree-Based Attribute Weighting): This type of method proposed by Hall [12] assigns small weight values to the attributes, which have strong dependencies on other attributes. In order to estimate each attribute's dependence on other attributes, an unpruned decision tree is constructed from the training instances with a minimum $depth$ indicating the depth for testing the tree. The weight for the attribute A_j is set as

$$w_j = \frac{1}{\sqrt{d_{A_j}}} \tag{10}$$

Where d_{A_j} is the minimum $depth$ at which the attribute A_j is tested in the tree. Attributes that do not appear in the tree receive a zero weight value.

III. EXPERIMENTS

A. Experimental Settings

We carry out the experiments for attribute weighted B-NCs using WEKA [19] data mining tool and validate their performance on 36 benchmark data sets from UCI data repository [13]. The data characteristics are described in [20]. Because Bayesian network classifiers are designed for categorical attributes, in our experiments, we first replace all missing attribute values using unsupervised attribute filter $ReplaceMissingValues$ in WEKA. Then, we apply unsupervised filter $Discretize$ in WEKA to discretize numeric attributes into categorical attributes.

In our experiments, all probability estimations use the Laplace Estimation, which introduces a prior probability for

TABLE IV

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED HNB VS. HNB WITH NO WEIGHT: ACC AND STANDARD DEVIATION.

Data Sets	HNB	CFS-HNB	GR-HNB	MI-HNB	Ref-HNB	DT-HNB
anneal	97.73±1.50	98.11±1.43	97.85±1.49	98.62±1.14 •	98.53±1.19	98.32±1.30 •
anneal.ORIG	88.92±3.02	89.37±2.91	89.41±2.81	91.60±2.63 •	90.51±2.96 •	90.13±2.99 •
audiology	75.87±6.30	75.41±5.82	76.22±6.35	73.15±6.00	74.70±6.12	76.39±6.47
autos	77.99±9.50	78.48±9.05	78.53±9.33	78.04±9.43	79.50±9.05	80.20±8.28
balance-scale	89.60±2.19	88.19±2.48 ○	89.41±2.30	89.65±2.42	88.82±2.49	89.35±2.31
breast-cancer	71.52±7.14	71.76±7.43	71.37±6.86	70.23±6.49	71.47±7.51	70.99±7.20
breast-w	96.32±2.47	96.32±2.38	96.37±2.46	96.08±2.46	95.88±2.56	95.95±2.47
colic	81.23±6.25	80.85±6.32	80.50±6.47	81.25±6.27	80.39±6.43	80.82±6.26
colic.ORIG	75.82±6.80	76.47±6.68	76.93±6.09	75.50±6.57	76.42±6.35	76.47±6.60
credit-a	85.35±3.88	85.12±3.76	84.49±3.72	84.84±4.43	84.77±4.09	85.12±3.89
credit-g	76.70±3.83	76.82±3.75	76.69±3.26	76.86±3.64	76.62±3.45	76.74±3.62
diabetes	75.13±4.91	74.87±4.89	74.70±4.90	75.83±4.86	74.41±4.77	75.30±4.97
glass	59.88±9.02	59.70±9.34	59.61±9.27	59.33±8.83	59.24±9.31	60.16±9.13
heart-c	82.38±6.95	82.68±6.94	82.05±7.09	81.43±7.35	82.02±7.08	82.48±6.80
heart-h	83.18±6.28	83.18±5.83	82.56±5.87	80.72±6.00	82.09±5.72	82.78±5.80
heart-statlog	82.33±5.91	82.63±5.61	82.59±5.48	81.74±5.94	82.37±6.10	82.44±5.64
hepatitis	84.18±9.50	84.37±9.37	84.75±8.75	82.71±9.95	84.44±9.46	84.05±9.66
hypothyroid	93.31±0.62	93.50±0.61	93.17±0.57	93.28±0.52	93.61±0.59	93.49±0.61
ionosphere	92.59±4.09	92.42±4.04	92.37±4.05	93.02±3.98	92.68±4.06	92.31±4.05
iris	93.00±6.02	91.40±5.94	92.47±5.58	93.93±6.00	92.20±6.22	90.87±6.47
kr-vs-kp	89.51±1.82	91.10±1.74 •	92.14±1.68 •	92.35±1.32 •	93.81±1.43 •	90.77±1.83 •
labor	93.37±11.35	93.20±11.37	92.87±11.41	90.87±13.15	92.13±12.56	91.40±12.52
letter	82.07±1.80	82.58±1.85 •	83.27±1.93 •	82.31±1.74	83.62±1.81 •	82.21±1.85
lymph	82.46±8.80	82.52±8.61	83.20±8.68	82.93±8.96	84.20±8.80	83.39±8.66
mushroom	99.88±0.26	99.88±0.26	99.88±0.26	99.94±0.19	99.88±0.26	99.81±0.34
primary-tumor	48.38±5.90	48.49±5.86	48.38±5.62	47.85±6.06	48.08±5.68	48.43±5.65
segment	94.19±1.38	94.26±1.36	94.41±1.37	94.72±1.42	94.77±1.31	94.21±1.34
sick	97.78±0.70	97.81±0.70	97.82±0.67	97.78±0.73	97.95±0.70	97.75±0.71
sonar	81.17±8.60	80.06±9.24	80.11±8.81	80.89±8.68	80.50±8.38	80.92±8.98
soybean	93.85±2.73	94.03±2.54	94.01±2.62	94.67±2.25	94.08±2.52	93.91±2.67
splice	96.06±1.02	96.08±1.08	96.06±1.07	96.13±0.99	96.09±1.06	96.10±1.06
vehicle	72.65±3.45	72.48±3.51	72.49±3.63	73.63±3.86	72.91±3.64	72.62±3.40
vote	93.56±3.45	93.24±3.41	93.72±3.20	94.36±3.20	92.02±3.73 ○	92.12±3.50 ○
vowel	92.03±2.58	91.79±2.65	91.00±2.77	92.99±2.49 •	91.79±2.66	91.92±2.63
waveform-5000	83.74±3.17	84.44±3.20	85.14±3.25	84.31±3.02	85.03±3.21	84.06±3.28
zoo	96.45±4.97	97.15±4.49	97.15±4.49	99.90±1.00 •	97.15±4.49	96.35±5.16
Mean	85.00±4.67	85.02±4.62	85.10±4.56	85.10±4.55	85.13±4.66	85.01±4.67
w/t/l	-	2/33/1	2/34/0	5/31/0	3/32/1	3/32/1

•, ○: Statistically significant upgradation and degradation, respectively.

each attribute A_j such that no attribute has a zero conditional probability values. In this case, The class probabilities $P(c_k)$, conditional probabilities $P(x_{t,i}, |c_k)$, and joint probabilities $P(x_{t,i}|x_{t,j}, c_k)$ in BNCs are estimated as

$$P(c_k) = \frac{F(c_k) + 1.0}{N + L} \tag{11}$$

$$P(x_{t,i}, |c_k) = \frac{F(x_{t,i}, c_k) + 1.0}{F(c_k) + |A_i|} \tag{12}$$

$$P(x_{t,i}|x_{t,j}, c_k) = \frac{F(x_{t,i}, x_{t,j}, c_k) + 1.0}{F(x_{t,i}, c_k) + |A_i|} \tag{13}$$

where, $|A_i|$ is the number of distinct values of attribute A_i and L is the number of classes in $\mathcal{D}$. $F(\cdot)$ is the frequency with which a combination of terms appears in the training data. N is the number of training instances. For MI-HNB, it employs the Eqs. (6) and (7) in [7] to calculate the $w_{i,j}$. While for other attribute weighted HNB models, the underlying $w_{i,j}$ can be obtained through $w_{i,j} = (w_i + w_j)/2$.

B. Evaluation Criterions

In our experiments, the selected algorithms are evaluated in terms of classification accuracy measured by ACC, ranking measured by AUC, and probability estimation measured by CLL. The ACC is calculated by the percentage of successful predictions on domain specific problems [21], [22], [23]. In addition to the accuracy, some data mining applications also require accurate rankings [24], so we also collect the AUC and CLL in our experiments, where AUC of the classifier is calculated as follows:

$$E = \frac{P_0 - t_0(t_0 + 1)/2}{t_0 t_1} \tag{14}$$

where t_0 and t_1 are the numbers of positive and negative instances, respectively. $P_0 = \sum r_i$, with r_i denoting the rank of ith positive instance in the ranked list. It is clear that AUC is essentially a measure of the quality of ranking, but it can only handle binary-class classification problems. For multiple classes, Hand and Till [25] propose another AUC measure:

$$E' = \frac{2}{L(L-1)} \sum_{i<j\leq L} E(c_i, c_j) \tag{15}$$

where L is the number of classes and $E(c_i, c_j)$ is the AUC of each pair of classes c_i and c_j.

CLL performance of a classfier h on data set $\mathcal{D}$ with N instances is evaluated by using Eq. (16).

$$CLL(h|\mathcal{D}) = \sum\nolimits_{t=1}^{N} logP_h(y_t|x_t) \tag{16}$$

where h is a learning model. In [3], maximizing Eq. (16) amounts to best approximation of the conditional probability of $\mathcal{Y}$ given each text instance x_t, and is equivalent to minimizing the conditional cross-entropy. The ACC, AUC, and CLL of each algorithm on each data set are obtained via 10 runs of 10-fold cross validation.

TABLE V

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED HNB VS. HNB WITH NO WEIGHT: AUC AND STANDARD DEVIATION.

Data Sets	HNB	CFS-HNB		GR-HNB		MI-HNB		Ref-HNB		DT-HNB	
anneal	99.04±1.43	99.07±1.41		99.06±1.42		99.15±1.37		99.19±1.36		99.12±1.40	
anneal.ORIG	98.04±2.41	97.98±2.65		97.99±2.69		97.87±3.27		98.02±2.72		98.06±2.54	
audiology	84.22±1.50	84.26±1.48		84.26±1.49		84.08±1.48		84.29±1.47		84.30±1.47	
autos	95.09±2.54	95.31±2.53		95.25±2.50		95.19±2.49		95.39±2.47		95.61±2.38	
balance-scale	87.21±4.13	85.98±4.37	○	86.99±4.24		87.38±4.21		86.75±4.33		87.03±4.28	
breast-cancer	69.52±10.06	68.86±10.16		66.40±10.18	○	66.69±10.61	○	69.41±9.87		69.24±10.17	
breast-w	99.10±0.89	99.09±0.91		99.11±0.89		99.02±0.95		99.08±0.92		99.09±0.90	
colic	86.35±6.25	86.28±6.21		86.26±6.08		86.73±6.09		86.21±6.17		86.57±6.22	
colic.ORIG	84.21±5.68	84.85±5.56		85.27±5.46		83.70±5.62		85.08±5.37		85.09±5.39	•
credit-a	91.81±3.11	91.85±3.02		91.89±2.95		91.04±3.58		91.76±3.01		91.76±3.10	
credit-g	79.56±4.12	79.44±4.15		79.08±4.14		79.65±4.42		79.09±4.15		79.48±4.12	
diabetes	82.34±4.93	81.98±4.86		81.59±4.88		82.31±4.82		81.47±4.82	○	82.19±4.88	
glass	88.58±4.33	88.74±4.30		88.68±4.37		88.37±4.57		88.67±4.47		88.66±4.37	
heart-c	84.03±0.62	84.00±0.63		83.99±0.63		83.94±0.63		83.97±0.63		84.00±0.63	
heart-h	83.93±0.57	83.94±0.55		83.92±0.55		83.79±0.59		83.91±0.54		83.90±0.56	
heart-statlog	90.13±5.02	89.88±5.08		89.79±5.19		89.26±5.22		89.61±5.23		89.85±5.10	
hepatitis	88.69±10.01	88.39±10.07		88.38±10.16		88.04±9.91		88.08±10.28		88.52±9.99	
hypothyroid	89.08±6.32	89.19±6.17		89.26±6.15		88.77±6.29		89.69±5.80		89.11±6.20	
ionosphere	97.90±1.99	97.63±2.12	○	97.67±2.10		98.19±1.95		97.93±1.99		97.25±2.60	○
iris	98.77±2.09	98.45±2.37		98.28±2.55		98.72±2.20		97.80±2.86		98.13±2.61	
kr-vs-kp	96.74±0.91	97.47±0.77	•	97.83±0.70	•	98.21±0.56	•	98.47±0.55	•	97.36±0.80	•
labor	97.33±7.34	97.33±7.12		97.50±6.98		97.04±7.52		97.04±7.52		97.71±6.92	
letter	98.86±0.21	98.91±0.21	•	98.99±0.19	•	98.89±0.20		99.03±0.19	•	98.88±0.21	•
lymph	95.01±4.24	95.06±4.24		94.93±4.23		94.82±4.27		95.24±4.19		95.11±4.29	
mushroom	100.00±0.01	100.00±0.01		100.00±0.01		100.00±0.01		100.00±0.01		100.00±0.01	
primary-tumor	85.84±2.10	86.00±2.05		85.92±2.06		85.86±2.12		86.10±2.06		85.87±2.10	
segment	99.59±0.18	99.62±0.16	•	99.67±0.15	•	99.71±0.14	•	99.71±0.14	•	99.61±0.17	
sick	97.16±1.73	97.77±1.43	•	98.07±1.36	•	98.24±1.22	•	98.84±0.86	•	98.18±1.25	•
sonar	90.03±6.57	90.11±6.58		90.01±6.71		90.15±6.63		90.55±6.35		89.99±6.84	
soybean	99.94±0.06	99.94±0.06		99.94±0.06		99.96±0.05		99.94±0.06		99.94±0.06	
splice	99.56±0.25	99.55±0.25		99.53±0.26		99.57±0.24		99.55±0.25		99.56±0.25	
vehicle	90.27±2.06	90.07±1.99		89.98±2.06		90.73±2.00	•	90.32±2.04		90.38±2.00	
vote	98.31±1.43	98.15±1.50	○	98.38±1.43		98.76±1.13	•	97.95±1.56	○	97.97±1.59	○
vowel	99.62±0.26	99.59±0.28		99.55±0.28	○	99.70±0.22	•	99.61±0.27		99.61±0.26	
waveform-5000	96.63±1.31	96.75±1.27		96.72±1.27		96.63±1.31		96.68±1.28		96.69±1.31	
zoo	99.07±1.43	99.07±1.43		99.07±1.43		99.26±1.11		99.07±1.43		99.07±1.43	
Mean	92.27±3.00	92.24±3.00		92.20±2.99		92.21±3.03		92.32±2.98		92.30±3.01	
w/t/l	-	4/29/3		4/30/2		6/29/1		4/30/2		4/30/2	

•, ○: Statistically significant upgradation and degradation, respectively.

C. Analysis of Attribute Weighted BNCs

Tables I, II, and III report the detailed results (the ACC, AUC, and CLL with the underlying standard deviation) of attribute weighted AODE, respectively. Besides, Tables IV, V, and VI report the detailed results of attribute weighted HNB, with the mean values and standard deviation on all data sets being summarized at the bottom, in which each entry $w/t/l$ means that the algorithm in the corresponding row wins in w data sets, ties in t data sets and loses in l data sets on the 36 benchmark data sets, compared to the BNC (AODE or HNB) without using attribute weighting. In these six tables, symbols • and ○ represent statistically significant upgradation and degradation over the BNC with their p-value less than 0.05 (*i.e.* a 95% confidence level). Overall, the results can be summarized as follows:

1. Attribute weighted AODE does not have significant superiority compared to AODE in ACC and AUC ranking. Specifically, attribute weighted AODE models slightly outperform AODE in ACC around (3 wins and 0 losses), and AUC around (4 wins and 0 losses).
2. Attribute weighted AODE significantly outperforms AODE in CLL, such as CFS-WAODE (10 wins and 1 losses), GR-WAODE (9 wins and 1 losses), MI-WAODE (8 wins and 1 losses), and Ref-WAODE (8 wins and 1 losses), except for DT-WAODE (7 wins and 2 losses) not so significant.
3. Attribute weighted HNB almost ties HNB with no weight in ACC around (2 wins and 0 losses), and AUC around (4 wins and 2 losses), except for that MI-HNB could slightly outperforms HNB in ACC (5 wins and 0 losses), and AUC (6 wins and 1 losses).
4. Attribute weighted HNB significantly outperforms HNB in CLL around (11 wins and 3 losses), except for DT-HNB (8 wins and 4 losses) with slight improvement.

Overall, the above observations suggest that attribute weighting could not achieve significant performance gain on BNCs (*e.g.* AODE and HNB) with complex structure, although the same weighting approaches have demonstrated good performance for the BNC (*e.g.* NB) with simple structure. One possible reason is that for complex BNCs, the structural dependency information in their Bayesian networks plays a very significant role in classification. So simply adding weight values to the attributes might not bring noticeable changes to the underlying classifiers. Detailed investigation in this direction is one of our future works in the next step.

IV. CONCLUSION AND FUTURE WORK

In this paper, we first reviewed the complex structure models for BNCs, and then carried out experimental studies to investigate the effectiveness of the attribute weighting strategies for complex BNCs, with a focus on two approaches, including hidden naive Bayes (HNB) and averaged one-dependence estimation (AODE). The experiments and com-

TABLE VI

EXPERIMENTAL RESULTS FOR ATTRIBUTE WEIGHTED HNB VS. HNB WITH NO WEIGHT: CLL AND STANDARD DEVIATION.

Data Sets	HNB	CFS-HNB	GR-HNB	MI-HNB	Ref-HNB	DT-HNB
anneal	-8.17±6.22	-7.15±6.14 •	-7.74±6.20 •	-5.67±6.22 •	-6.14±6.35 •	-6.65±6.14 •
anneal.ORIG	-23.14±6.71	-22.27±6.65 •	-21.76±6.78 •	-19.44±6.91 •	-20.64±6.96 •	-21.16±6.76 •
audiology	-232.55±115.09	-234.23±115.35	-227.61±115.44 •	-285.76±114.34 ○	-245.34±116.25 ○	-232.71±116.64
autos	-155.99±82.37	-151.51±83.72	-149.80±80.00 •	-157.88±86.30	-149.19±82.31 •	-144.87±81.12 •
balance-scale	-45.86±3.27	-47.12±3.48 ○	-46.07±3.31 ○	-45.99±3.51	-46.24±3.42 ○	-46.09±3.30 ○
breast-cancer	-61.68±12.54	-61.43±11.72	-62.24±10.60	-63.75±12.14	-62.83±13.13	-61.65±12.20
breast-w	-14.07±8.87	-14.52±8.95	-14.06±8.80	-14.76±8.82	-14.74±9.04	-14.56±8.92
colic	-60.08±21.15	-59.25±20.50	-59.81±20.23	-57.17±19.77 •	-60.65±20.82	-59.32±21.07
colic.ORIG	-53.27±13.83	-51.91±13.34 •	-51.19±13.17	-55.90±14.64	-52.14±13.78	-51.79±13.21
credit-a	-39.80±10.51	-38.81±9.91 •	-37.82±9.01	-40.59±10.76	-38.53±9.41	-39.49±10.27
credit-g	-51.54±6.11	-51.45±6.00	-51.90±6.00	-51.30±6.34	-52.04±6.12	-51.55±6.07
diabetes	-52.46±9.34	-53.20±9.45	-54.17±9.76 ○	-52.81±9.41	-54.29±9.59 ○	-52.79±9.41
glass	-100.98±25.97	-101.31±26.60	-101.09±26.48	-101.93±25.68	-102.39±26.80	-100.98±26.29
heart-c	-46.17±21.54	-46.14±21.28	-46.03±21.05	-47.27±20.79	-46.84±21.50	-46.26±21.35
heart-h	-44.69±16.31	-43.17±14.76	-42.09±13.33	-46.86±15.18	-42.89±13.17	-43.95±14.68
heart-statlog	-48.63±18.79	-48.62±18.48	-48.14±18.01	-50.69±18.58	-49.38±18.76	-48.95±18.61
hepatitis	-45.24±29.35	-44.60±28.57	-44.16±28.20	-44.41±25.48	-44.22±27.66	-44.38±28.44
hypothyroid	-23.59±2.94	-23.17±2.82 •	-24.08±2.94	-23.05±2.71	-22.47±2.76 •	-23.19±2.84 •
ionosphere	-54.74±39.24	-56.96±39.64	-56.24±39.01	-51.28±37.22	-54.18±39.49	-61.59±43.67 ○
iris	-20.74±14.94	-22.34±15.63 ○	-23.44±16.17	-19.98±15.31	-27.31±18.94 ○	-24.79±17.04 ○
kr-vs-kp	-26.18±2.04	-24.41±1.93 •	-22.84±1.91 •	-21.82±1.76 •	-20.19±1.84 •	-24.46±1.98 •
labor	-20.85±28.71	-21.33±29.33	-20.84±28.12	-27.27±35.73 ○	-26.35±36.01	-24.12±33.62
letter	-66.34±6.77	-64.37±6.71 •	-62.08±6.54 •	-65.04±6.55 •	-60.97±6.42 •	-65.74±6.79 •
lymph	-42.11±23.71	-42.70±23.18	-42.68±22.88	-45.16±24.88	-41.38±22.90	-42.29±22.97
mushroom	-0.48±1.03	-0.49±0.93	-0.44±0.77	-0.38±0.91	-0.40±0.85	-0.88±1.49
primary-tumor	-194.35±25.64	-193.67±25.34	-193.81±25.46	-194.52±25.23	-193.90±25.40	-194.15±25.54
segment	-16.94±4.90	-16.26±4.63 •	-14.99±4.32 •	-14.27±4.28 •	-14.67±4.47 •	-16.77±4.82
sick	-8.72±2.42	-7.61±2.11 •	-7.16±2.08 •	-6.80±1.96 •	-5.81±1.62 •	-6.98±1.95 •
sonar	-75.75±41.50	-73.23±39.52	-73.48±39.85	-76.33±41.07	-76.03±40.70	-75.77±43.64
soybean	-25.11±10.73	-22.50±9.60 •	-24.64±10.48	-13.22±5.29 •	-20.59±8.71 •	-25.03±10.81
splice	-12.38±3.55	-12.29±3.43	-12.35±3.39	-11.94±3.31 •	-12.26±3.42	-12.27±3.44
vehicle	-72.57±10.29	-70.93±9.83 •	-68.89±9.49 •	-63.34±8.95 •	-68.48±9.67 •	-72.31±10.21
vote	-25.16±14.87	-23.83±13.29	-19.42±10.91 •	-17.30±10.74 •	-26.56±14.20	-27.21±14.88 ○
vowel	-23.33±7.54	-24.28±7.65 ○	-25.98±8.07 ○	-20.34±7.31 •	-24.26±7.79	-23.52±7.57
waveform-5000	-47.27±10.87	-40.97±9.56 •	-36.12±8.63 •	-42.23±9.89 •	-36.44±8.64 •	-43.76±10.12 •
zoo	-7.22±7.94	-7.13±7.68	-7.57±8.47	-4.21±3.90	-7.39±8.31	-7.83±8.50
Mean	-51.34±18.54	-50.70±18.27	-50.08±17.94	-51.69±18.11	-50.78±18.53	-51.11±18.79
w/t/l	-	12/21/3	11/22/3	13/21/2	11/21/4	8/24/4

•, ○: Statistically significant upgradation and degradation, respectively.

parisons on 36 benchmark data sets with respect to the classification accuracy, class probability estimation, and ranking performance showed that although attribute weighting can significantly improve NB classifier with simple structure, the improvement of attribute weighting for general BNCs with complex structures is, nevertheless, insignificant.

REFERENCES

[1] J. Hernández-González, I. n. Inza, and J. A. Lozano, "Learning bayesian network classifiers from label proportions," *Pattern Recogn.*, vol. 46, no. 12, pp. 3425–3440, 2013.

[2] P. Langley and S. Sage, "Induction of selective bayesian classifiers," in *UAI*, 1994, pp. 339–406.

[3] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2-3, pp. 131–163, 1997.

[4] R. Kohavi, "Scaling up the accuracy of naive-bayes classifiers:a decision-tree hybrid," in *KDD*, 1996, pp. 202–207.

[5] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so naive bayes: Aggregating one-dependence estimators," *Machine Learning*, vol. 58, pp. 5–24, 2005.

[6] J. Su and H. Zhang, "Full bayesian network classifiers," in *ICML*, 2006, pp. 897–904.

[7] L. Jiang, H. Zhang, and Z. Cai, "A novel bayes model: Hidden naive bayes," *IEEE Trans. on Knowl. and Data Eng.*, vol. 21, no. 10, pp. 1361–1371, 2009.

[8] L. Jiang, H. Zhang, Z. Cai, and D. Wang, "Weighted average of one-dependence estimators?." *J. Exp. Theor. Artif. Intell.*, vol. 24, pp. 219–230, 2012.

[9] J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[10] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *ICML*, 2000, pp. 359–366.

[11] I. Kononenko, "Estimating attributes: analysis and extensions of relief," in *ECML*, Secaucus, NJ, USA, 1994, pp. 171–182.

[12] M. Hall, "A decision tree-based attribute weighting filter for naive bayes," *Know.-Based Syst.*, vol. 20, no. 2, pp. 120–126, 2007.

[13] K. Bache and M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[14] C. X. Ling, J. Huang, and H. Zhang, "Auc: A statistically consistent and more discriminating measure than accuracy," in *IJCAI*, 2003, pp. 519–524.

[15] D. Grossman and P. Domingos, "Learning bayesian network classifiers by maximizing conditional likelihood," in *ICML*, 2004, pp. 361–368.

[16] N. A. Zaidi, J. Cerquides, M. J. Carman, and G. I. Webb, "Alleviating naive bayes attribute independence assumption by attribute weighting," *Journal of Machine Learning Research*, vol. 14, pp. 1947–1988, 2013.

[17] H. Zhang and S. Sheng, "Learning weighted naive bayes with accurate ranking," in *ICDM*, 2004, pp. 567–570.

[18] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, San Francisco, CA, USA, 1992, pp. 249–256.

[19] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., ser. The Morgan Kaufmann Series in Data Management Systems. San Francisco, CA: Morgan Kaufmann Publishers, 2005. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

[20] J. Wu, Z. Cai, and X. Zhu, "Self-adaptive probability estimation for naive bayes classification," in *IJCNN*, 2013, pp. 2303–2310.

[21] J. Wu, X. Zhu, C. Zhang, and Z. Cai, "Multi-instance multi-graph dual embedding learning," in *ICDM*, 2013, pp. 827–836.

[22] S. Pan, X. Zhu, C. Zhang, and P. Yu, "Graph stream classification using labeled and unlabeled graphs," in *ICDE*, 2013, pp. 398–409.

[23] J. Wu, X. Zhu, C. Zhang, and P. Yu, "Bag constrained structure pattern mining for multi-graph classification," *Knowledge and Data Engineering, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[24] H. Zhang and J. Su, "Naive bayesian classifiers for ranking," in *ECML*, 2004, pp. 501–512.

[25] D. J. Hand and R. J. Till, "A simple generalisation of the area under the roc curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.