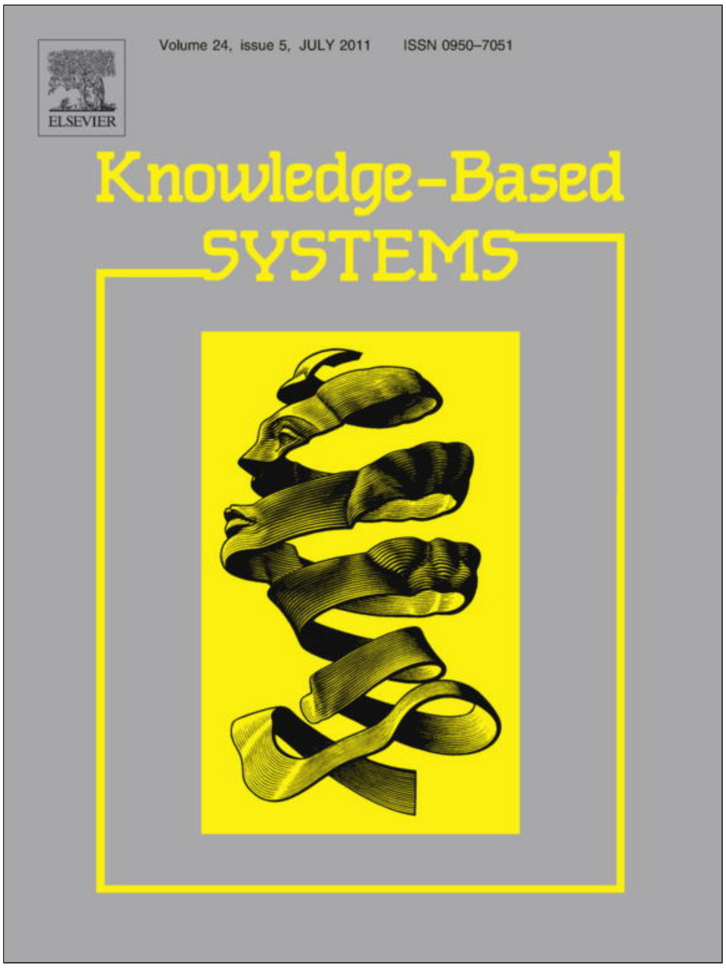
Volume 24, issue 5, JULY 2011 ISSN 0950-7051
ELSEVIER
Knowledge-Based
SYSTEMS

Knowledge-Based Systems 24 (2011) 563–569

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

NB$^+$: An improved Naïve Bayesian algorithm

Appavu alias Balamurugan *, Ramasamy Rajaram, S. Pramala, S. Rajalakshmi, C. Jeyendran, J. Dinesh Surya Prakash

Department of Information Technology, Thiagarajar College of Engineering, Madurai 15, India

ARTICLE INFO

Article history:
Received 17 June 2008
Received in revised form 26 September 2010
Accepted 26 September 2010
Available online 7 October 2010

Keywords:
Data mining
Classification
Naïve Bayesian algorithm
Conflicting data
Lazy learning

ABSTRACT

A novel algorithm named NB$^+$ which is an extended version of the traditional Naïve Bayesian algorithm has been presented in this paper. An exception occurs when there is an equal probability for the class label value in the Naïve Bayesian algorithm. The approach aims to suggest a solution with the help of a partial matching method. Consequently, the classification accuracy has drastically improved. Experimental evaluation has been done on various databases to show that NB$^+$ algorithm outperforms the traditional Naïve Bayesian algorithm.

1. Introduction

In data mining we sort, extract and establish relationships between huge amounts of data. Potentially useful information from large datasets or databases can be extracted. We can relate data mining and enterprise resource planning (ERP) to identify patterns during the analysis of the large sets of transaction data.

Classification in data mining is a procedure where grouping of information is done based on certain traits or characteristics. In the learning step or the training phase of a classification algorithm, analysis of the training set is done to build the classifier $Y = f(X)$, which can be called as a mapping function can predict the associated class label Y for a given tuple X. This can represented as either rules or decision tress or formulae.

In the second step, while measuring the predictive accuracy of the classifier we use the training set in order to get an optimistic estimate. A test set consisting of test tuples and their corresponding class labels is used. These randomly selected tuples are independent of the training tuples and are not used in the model generation. If the accuracy of the classifier is considered acceptable, it can then be used to classify unknown sample. There are many types of classifiers like Bayesian classification, classification by decision tree induction, and back propagation.

Since the Naïve Bayes algorithm's conditional independence assumption is rarely true, efforts are being made to improve the efficiency of this algorithm. Other aspects of knowledge discovery, such as identification of relevant features and inconsistent data are only of secondary importance to the Naïve Bayes algorithm. It also cannot classify a particular kind of record when the training data satisfy the constraint given below:

1. The probability of every class label attribute is evenly distributed among the distinct attribute values.

The NB$^+$ algorithm presented in this paper aims at deciding the correct class label value when the training data have the above mentioned constraints. It is important to solve this problem because there are several real world datasets where the classification of the given unknown sample has to be accurate. For example in a dataset belonging to a hospital repository with various symptoms as their attributes and final diagnosis as the class value, it is very essential that a more precise classification is done to get a correct diagnosis. If we get a sample from a patient having symptoms that do not exactly match any tuple in our dataset, then this 'closely matching attribute selection' method aids in diagnosing the disease more accurately, rather than a random pick as in the traditional Naïve Bayes algorithm. This paper conducts a series of tests on the real world data to evaluate the approach. The approach described is general and can be used to analyze not only real world data but also other types of databases. The paper is organized in the following manner: Section 2 defines related works in this area. Section 3 portrays the problem handled in this paper. Section 4 explains our proposed algorithm with experimental examples. Section 5 gives a note of comparison between the Naïve Bayes classifier and the proposed algorithm, highlighting its advantages. Finally, Section 6 summarizes the proposed algorithm and concludes the paper.

* Corresponding author.
E-mail address: app_s@yahoo.com (Appavu alias Balamurugan).

0950-7051/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2010.09.007

2. Related works

The Naïve Bayes classifier (NB) has been a widely researched classification model, used in many cases as a benchmark for comparison to new methods. The elegant simplicity and apparent accuracy of NB even when the independence assumption is violated [6] fosters the on going interest in the model. Much of the current research is based on the idea that "small alterations" to NB in the literature is growing fast. As there is an abundance of proposed adjustments to NB in the literature to find the model with the correct adjustments for a particular type of data would mean an empirical analysis of many models. This would lead to a result only applicable to that data. This study uses a form of meta-analysis to give a broader result. Meta-analysis is a technique for comparing different studies to try and draw a concensus of option [50]. It is used mainly in the social sciences, biology and psychology, [13,23]. Meta-analysis has been used rarely in pattern recognition with a meta-analysis of classification-algorithms given by Lu et al. [40], a meta-analysis of face recognition algorithms given by Phillips and Newton [30] and most recently a landscape of clustering algorithms given by Chen et al. [14]. The idea behind Bayesian networks is given in [29]. A unique joint probability distribution over the set of features is defined by the network.

Boosting is a machine learning technique that focuses the learning of a model on the cases that are hard to classify. The first polynomial-time boosting algorithm was proposed by Schapire [37]. This was improved upon by Freund [9] to make it more efficient. This improvement was optimal in many cases but had practical drawbacks. Adaboost proposed by Freund [10] solved many of the practical difficulties of boosting and is widely used today. An introduction to boosting algorithms is given by Schapire [38]. A comparison study by Lewis [22] concentrated in applying NB to textual data. The literature of the Naïve Bayes algorithm has been well-studied in [7,19,25,36,39]. The various approaches proposed in the literature have been reviewed to improve the performance of the Naïve Bayes algorithm.

In a decision tree Naïve Bayes hybrid (NBTree) [18], a standard decision tree is grown with a NB deployed at the leaves, creating a penultimate layer of the tree. These classifiers leave a classifier label as the output when a new case is submitted to them, acting as the final decision of the tree. Iterative Bayes (IB), [11] begins with a contingency table built by the standard NB. An iterative procedure then updates these tables by cycling through all the training examples. In ensemble feature selection Naïve Bayes (EFSNB) [44], an ensemble of NB classifiers is created. Randomly sampled subset of the original set of features is used to train each NB. In each of the following methods, sequential forward selection Naïve Bayes (SFSNB), sequential backward selection Naïve Bayes (SBSNB), genetic algorithm Naïve Bayes (GANB), [43] an ensemble of NB classifiers is created. Each NB is trained on a subset of the original features. The selection of the feature subsets is achieved in a sequential forward selection, in a sequential backward selection and by genetic algorithm methods, respectively. Xiao et al. [15] introduces group method of data handling theory to Bayesian classification and proposed GMBC algorithm for structure identification of Bayesian classifiers. On the classification performance of TAN and general Bayesian networks given by [26].

In a tree augmented Bayesian network (TAN), [8] a network is grown in which none of the class labels have a parent. For every feature the class variable is assigned as its parent and at most one other feature, making TAN a one-dependency network. In the probability dependence Tree2 (PDT2), [16] the network is initialized to NB. Each node is considered for the super parent in turns by extending edges to every node which does not have a parent, besides the class node. In sequential forward selection and Naïve Bayes (SFS to Bayes), [21] a subset of the original feature is chosen by sequential forward selection. A NB is then trained using only these selected features. Using sequential forward selection and joining (SFSJ), and sequential backward selection and joining (SBSJ), [28] the subsets of features are selected for the NB classifier. At each step of the feature selection, one of the features may be added to (SFS), or removed from (SBS) the subset, or a feature may be joined to another one already present in the subset. In a K-Dependence Bayesian network (KDEPBN), [35] the space of k dependencies is searched for the most appropriate Bayesian network for the problem. The value of k is decided by the user. Aggregating one dependence estimators (AODE), [48] the class-conditional pdf, $P(x|w_k)$ is approximated as the average of n "mini"-pdfs, one for each feature. In lazy Bayesian rules (LBR), [52] each test case LBR generates an appropriate rule with a conjunction of feature-value pairs as its antecedent and a local NB as its consequent. The conditional independence tree (CI-Tree), [51] represents a joint distribution over all the features explicitly defining the conditional dependencies among them. Chen et al. [14] proposed a selective Bayes classifier for classifying incomplete data based on gain ratio.

Independent component analysis Naïve Bayes (ICABayes), [2] the independent analysis of an n-dimensional random vector is the linear transform which minimises the statistical dependence between its components. The lazy version of the tree augmented network (Lazy-TAN), [47] is the lazy variant of the super parent algorithm PDT2, [16]. The interval estimation Naïve Bayes (IENB), [33] calculates confidence intervals for the NB point estimations of $P(x_{i|}w_j)$. In a random tree augmented network (RTAN), [24] an ensemble of TAN classifiers is grown. It is then integrated using the majority voting method.

In the adapted boosting for Naïve Bayes (Active Boost), [46] a new test case is labelled and then added to the training set. The updated data set is then used to train another NB. Adjusted probability Naïve Bayes classifier [49] uses the probability distributions produced by NB. Homologous Naïve Bayes (HNB), [12] takes advantage of the knowledge that multiple cases submitted for labeling come from the same unknown class. In fuzzy Naïve Bayes (FNB), [41] each feature value of x is accompanied by a degree of membership in the interval $[0,1]$. The interpretable boosted Naïve Bayes (IBNB), [32] method aims to improve NB by boosting. Yet it still has an end product that is interpretable by the user. The semi-Naïve Bayes (SNB), [20] partitions the feature into groups using statistical tests of independence. Kernel-based and joining Naïve Bayes (KJNB), [4] features are joined if they are highly correlated. In Naïve Bayes committees (NBC), [53] a set of NB classifiers are generated in sequential trials. NB base is generated as the founder of the 'committee' using all of the features.

In boosted Naïve Bayes (BNB), [5] the boosting strategy is applied to NB. The training samples are selected by the bootstrap method. In clustered Naïve Bayes (CNB), [45] the examples from each class are clustered. The training data is then relabeled using the cluster labels. The neuro-fuzzy Naïve Bayes (NFNB), [27] derives fuzzy classifiers from data using neural-network inspired learning. The method maps NB to a neuro-fuzzy classifier. Minimum Description Length (MDL) principle in Naïve Bayes (MNB), [17] starts with a Bayesian network representing class-conditional independence. Selective Bayesian classifier (SBC), [31] runs the decision tree algorithm, C 4.5 on 10% of the training set. The features on the first three levels of the decision tree are selected. This is repeated five times on different 10% selections of the training data. In boosted levelled Naïve Bayes trees (BLNBT), [42] a standard decision tree of user defined depth is grown as in the Nbtree [18] method. The extended Bayes (EB), [34] finds sets of dependent features using information gain measure. [1] Presented a novel algorithm named ID6NB for extending decision tree induced by

Quinlan's non-incremental ID3 algorithm. Chinese text classification by the Naïve Bayes Classifier and the associative classifier with multiple confidence threshold values given in [40].

3. Problem definition

The constraint due to which both the Naïve Bayes algorithm fail to classify a particular kind of instance is given below:

"The probability of every class attribute is evenly distributed among the distinct attribute values".

In the case of Naïve Bayes algorithm, incorrect and inaccurate classification occurs due to the random assignment of class labels.

3.1. Resolving problem in the Naïve Bayes

The Naïve Bayes algorithm fails when "The probability of every class label attribute is evenly distributed among the distinct attribute values in the training dataset".

Here the condition $P(X|C_i)\ P(C_i) = P(X|C_j)\ P(C_j)$ for $1 \leqslant j \leqslant m, j \neq i$, where m refers to number of classes $C_1, C_2 \ldots C_m$

3.1.1. Illustrative example on a real world dataset

Consider Table 1 which contains the dataset possessing the constraints mentioned in the problem definition.

The data samples are described by the attributes BI-RADS, Age, Shape, Margin and Density. The class label attribute, Severity, has two distinct values (namely, {malignant, benign}). Let C_1 correspond to the class Severity = "malignant" and C_2 correspond to the class Severity = "benign". The unknown sample we wish to classify is X = (BI-RADS = "E", Age = "Senior", Shape = "Irregular", Margin = "Circumscribed", Density = "iso")

P (Severity = "malignant") = 1/2
P (Severity = "benign") = 1/2
P (BI-RADS = "E" | severity = "malignant") = 1/2
P (BI-RADS = "E" | severity = "benign") = 1/2
P (Age = "senior" | severity = "malignant") = 1/2
P (Age = "senior" | severity = "benign") = 1/2
P (shape = "irregular" | severity = "malignant") = 1/2
P (shape = "irregular" | severity = "benign") = 1/2
P (margin = "circumscribed" | severity = "malignant") = 1/2
P (margin = "circumscribed" | severity = "benign") = 1/2
P (density = "iso" | severity = "malignant") = 1/2
P (density = "iso" | severity = "benign") = 1/2

Table 1
Mammographic mass dataset.

BI-RADS	Age	Shape	Margin	Density	Severity
C	M	Round	Obscured	High	Benign
E	Y	Irregular	Spiculated	High	Malignant
C	Y	Irregular	Spiculated	Low	Benign
A	M	Lobular	Ill-defined	Low	Malignant
A	Y	Round	Ill-defined	High	Malignant
E	S	Irregular	Spiculated	Iso	Malignant
A	S	Irregular	Circumscribed	Iso	Benign
D	S	Oval	Circumscribed	High	Benign
A	Y	Oval	Spiculated	Low	Benign
B	S	Oval	Obscured	Low	Malignant
B	Y	Oval	Obscured	Iso	Malignant
E	S	Irregular	Obscured	Iso	Benign
D	S	Irregular	Circumscribed	Fat	Malignant
C	Y	Lobular	Microlobulated	High	Malignant
B	Y	Irregular	Ill-defined	Fat	Benign
C	M	Round	Microlobulated	High	Malignant
D	S	Irregular	Circumscribed	Iso	Malignant
E	S	Lobular	Microlobulated	Iso	Benign
D	Y	Round	Microlobulated	High	Benign
B	M	Lobular	Ill-defined	High	Benign

Using the above probabilities we obtain.

$P(X$ | severity = "malignant") = 1/2 ∗ 1/2 ∗ 1/2 ∗ 1/2 ∗ 1/2 = 1/32
$P(X$ | severity = "benign") = 1/2 ∗ 1/2 ∗ 1/2 ∗ 1/2 ∗ 1/2 = 1/32
$P(X$ | severity = "malignant") $P(X$ | severity = "malignant") = /32 ∗ 1/2 = 0.0156
$P(X$ | severity = "benign") $P(X$ | severity = "benign") = 1/32 ∗ 1/2 = 0.0156.

From this example it is clear that the Naïve Bayes algorithm fails when

$$P(X|C_i)P(C_i) = P(X|C_j)P(C_j)$$

Since there is equal probability for Severity = "malignant" and Severity = "benign", we are in a fix. This is the problem found in the Naïve Bayes algorithm. In order to resolve the problem the improved Naïve Bayes is introduced.

4. The proposed learning algorithm

NB update procedure to handle the cases where the class probabilities are exactly the same is given below.

1. Let $A_1 \ldots A_n$ be the total number of predictive Attributes in both the test and training data where n refers to maximum number of predictive Attributes.
2. Let $C_1 \ldots C_m$ be the different class labels assigned to the training data where m refers to the number of classes.
3. The Influence Factor is calculated for the training data. The formula for the Influence Factor is given below:

   $$\text{Influence Factor } I(X|C_i) = \frac{N(X|C_i)}{N(C_i)}$$

   where $N(X | C_i)$ = Number of records in which attribute value X has the class label C_i.
   $N(C_i)$ = Total Number of records in which the class label is C_i.
   The Influencing Factor gives the dependability of the attribute value on the class attribute. The Steps involved are:
   (i) The records having the same class label values are extracted from the training dataset and made into separate sub tables. If there are 'm' lass labels then we have 'm' numbers of sub tables with 'i' records in each sub table.
   (ii) The Influence Factor is found for all the attribute values. The values with highest Influence Factor are only considered and their corresponding attributes are called Influencing Attributes. Other attributes are called Non Influencing Attributes.
4. The test data is compared with every record in the training set and let 'q' be the maximum number of attributes that match. As $q < n$, many different combinations of q attribute sets are possible which are called 'Available Attribute Combinations' (AAC).
5. If there is only one AAC then
   (i) If all the records inside this AAC have a class label say C_i then that class label is directly assigned to the test data.
   (ii) If the records inside this AAC have different class labels say $C_i, C_j, C_k \ldots$ then the occurrence of each of the class label is found.
      a. If there is equal occurrence for each of the class label then we go for the next AAC with $q - 1$ attributes. For e.g. consider an AAC with class labels C_i and C_j. If both the class labels C_i and C_j occur twice in the AAC then we can't decide upon the class label and hence we go for the next AAC with $q - 1$ attributes.

b. If the occurrence of each one of the class labels in the AAC is different, then the one with the greatest occurrence is selected as the class label for that AAC and the same is assigned to the test data.

6. If there two or more AACs, then the occurrence of the class label in the records of each one of the AACs is found. If there is equal occurrence for every class label, such AAC's are called EAAC i.e., 'Equal occurrence in Available Attribute Combination' else it is called NAAC i.e., 'Non-Equal occurrence in Available Attribute Combination'. The EAACs are dropped and they are not used to classify the test data. If all the obtained AACs are EAACs, we go for the next iteration.
7. If an AAC is NAAC, the one with the greatest occurrence is selected as the class label for that particular AAC. Similarly the class labels for all the NAACs' are found.
   (a) If all the NAACs result in the same class label, that class label is assigned to the test data.
   (b) If each NAAC results in different class label, the above mentioned Step 3 is executed i.e., the Influence Factor is found for all the NAACs separately. The NAACs' with the highest number of Influencing Attributes are considered for the classification.
      (i) If there is only one NAAC with highest number of Influencing Attributes, its class label is directly assigned to the test data.
      (ii) If there are two or more NAACs with highest number of Influencing attributes then
   a. If the class label for all these NAACs are the same, it is directly assigned to the test data.
   b. If the class label is different for different NAACs, the occurrence of each of the class label is found.
   - If there is equal occurrence for each of the class label, we go for the next AAC with $q-1$ attributes.
   - If the occurrence of each of the class label in the NAAC is different, the one with the greatest occurrence is selected as the class label for that NAAC and the same is assigned to the test data.

5. Implementation of the proposed algorithm

To prove the efficiency of the proposed algorithm we consider Table 1 used in the problem definition. When the probability of every class label is evenly distributed among the distinct attribute values, exceptions occur in the Naïve Bayes algorithm.

In the case of NB the class label values are equal. So an updated procedure has been proposed to decide upon the class label. The unknown sample we wish to classify is X = (BI-RADS = "E", Age = "Senior", Shape = "Irregular", Margin = "Circumscribed", Density = "iso").

5.1. Finding influence factor

Step 1: Divide the training data based on the class label. Here in this example the records having the class label 'malignant' are separated from the records having the class label value 'benign' and so two tables are got. Training data having class label malignant and benign are shown in Table 2a and 2b.

Step 2: Find the influence factor for all the attribute values. The influence factor gives the dependability of the attribute value on the class label.

$$\text{Influence Factor } I(X|C_i) = \frac{N(X|C_i)}{N(C_i)}$$

where $N(X|C_i)$ = Number of records in which attribute value X has the class label C_i.$N(C_i)$ = Total Number of records in which the class label is C_i.

Table 2a
Training data having class label malignant.

BI-RADS	Age	Shape	Margin	Density	Severity
E	Y	Irregular	Spiculated	High	Malignant
A	M	Lobular	Ill-defined	Low	Malignant
A	Y	Round	Ill-defined	High	Malignant
E	S	Irregular	Spiculated	Iso	Malignant
B	S	Oval	Obscured	Low	Malignant
B	Y	Oval	Obscured	Iso	Malignant
D	S	Irregular	Circumscribed	Fat	Malignant
C	Y	Lobular	Microlobulated	High	Malignant
C	M	Round	Microlobulated	High	Malignant
D	S	Irregular	Circumscribed	Iso	Malignant

Table 2b
Training data having class label benign.

BI-RADS	Age	Shape	Margin	Density	Severity
C	M	Round	Obscured	High	Benign
C	Y	Irregular	Spiculated	Low	Benign
A	S	Irregular	Circumscribed	Iso	Benign
D	S	Oval	Circumscribed	High	Benign
A	Y	Oval	Spiculated	Low	Benign
E	S	Irregular	Obscured	Iso	Benign
B	Y	Irregular	Ill-defined	Fat	Benign
E	S	Lobular	Microlobulated	Iso	Benign
D	Y	Round	Microlobulated	High	Benign
B	M	Lobular	Ill-defined	High	Benign

Step 2a: Find the influence factor for all the attribute values of the Table 2a.

BI-RADS Attribute:
I (BI-RADS = "A" | severity = "malignant") = 2/10
I (BI-RADS = "B" | severity = "malignant") = 2/10
I (BI-RADS = "C" | severity = "malignant") = 2/10
I (BI-RADS = "D" | severity = "malignant") = 2/10
I (BI-RADS = "E" | severity = "malignant") = 2/10
The impact factor is equal for all the attributes. So the attribute 'BI-RADS' does not influence the class label.

AGE Attribute:
I(age = "senior" | severity = "malignant") = 4/10
I (age = "medium" | severity = "malignant") = 2/10
I (age = "young" | severity = "malignant") = 4/10
The impact factor is more for the attribute values senior and young. So the values 'S' and 'Y' of the 'AGE' attribute have greater influence on the class label.

SHAPE Attribute:
I (shape = "irregular" | severity = "malignant") = 4/10
I (shape = "oval" | severity = "malignant") = 2/10
I (shape = "lobular" | severity = "malignant") = 2/10
I (shape = "round" | severity = "malignant") = 2/10
The impact factor is more for the attribute value irregular. The value 'Irregular' of the 'SHAPE' attribute influences the class label more.

MARGIN Attribute:
I(margin = "circumscribed" | severity = "malignant") = 2/10
I (margin = "spiculated" | severity = "malignant") = 2/10
I (margin = "obscured" | severity = "malignant") = 2/10
I (margin = "ill-defined" | severity = "malignant") = 2/10
I (margin = "microlobulated" | severity = "malignant") = 2/10
The impact factor is equal for all the attributes. So the attribute 'MARGIN' does not influence the class label.

DENSITY Attribute:
I (density = "iso" | severity = "malignant") = 3/10
I (density = "fat" | severity = "malignant") = 1/10

I (density = "low" | severity = "malignant") = 2/10
I (density = "high" | severity = "malignant") = 4/10
The impact factor is more for the attribute value 'high'. Its influence on the class label is more.

Step 2b: Find the influence factor for all the attribute values of the Table 2b.

BI-RADS Attribute:
I (BI-RADS = "A" | severity = "benign") = 2/10
I (BI-RADS = "B" | severity = "benign") = 2/10
I (BI-RADS = "C" | severity = "benign") = 2/10
I (BI-RADS = "D" | severity = "benign") = 2/10
I (BI-RADS = "E" | severity = "benign") = 2/10
The impact factor is equal for all the attributes. So none of the attribute values influence the class label.

AGE Attribute:
I (age = "senior" | severity = "benign") = 4/10
I (age = "medium" | severity = "benign") = 2/10
I (age = "young" | severity = "benign") = 4/10
The impact factor is more for the attribute values senior and young. So the values 'S' and 'Y' of the 'AGE' attribute have greater influence on the class label.

SHAPE Attribute:
I (shape = "irregular" | severity = "benign") = 4/10
I (shape = "oval" | severity = "benign") = 2/10
I (shape = "lobular" | severity = "benign") = 2/10
I (shape = "round" | severity = "benign") = 2/10
The impact factor is more for the attribute value 'irregular'. It has greater influence on the class label.

MARGIN Attribute:
I (margin = "circumscribed" | severity = "benign") = 2/10
I (margin = "spiculated" | severity = "benign") = 2/10
I (margin = "obscured" | severity = "benign") = 2/10
I (margin = "ill-defined" | severity = "benign") = 2/10
I (margin = "microlobulated" | severity = "benign") = 2/10
The impact factor is equal for all the attributes. So none of the attribute values influence the class label.

DENSITY Attribute:
I (density = "iso" | severity = "benign") = 3/10
I (density = "fat" | severity = "benign") = 1/10
I (density = "low" | severity = "benign") = 2/10
I (density = "high" | severity = "benign") = 4/10
The impact factor is more for the attribute value high. It influences the class label more.

5.2. Applying updated procedure on the training set

Step 1: We find the tuples which closely match the unknown sample. They are shown in Table 3.

Step 2: We now pick out the closely matching attributes from these closely matching tuples. They are shown in Table 4. The second set of closely matching attributes from Table 4 is shown in Table 5.

Step 3: We now find that the class label values 'malignant' and 'benign' are present in equal probabilities in both the 4-attribute sets. A tie occurs and we are not able to decide upon the class label value.

Table 3
Closely matching tuple from Table 1.

BI-RADS	Age	Shape	Margin	Density	Severity
E	S	Irregular	Spiculated	Iso	Malignant
E	S	Irregular	Obscured	Iso	Benign
D	S	Irregular	Circumscribed	Iso	Malignant
A	S	Irregular	Circumscribed	Iso	Benign

Table 4
First set of closely matching attributes from Table 3.

BI-RADS	Age	Shape	Margin	Density	Severity
E	S	Irregular	Spiculated	Iso	Malignant
E	S	Irregular	Obscured	Iso	Benign

In the first one it is BI-RADS, Age, Shape and Density and in the second one it is Age, Shape, Margin and Density. We take four closely matching attributes here. Selected attributes: BI-RADS, Age, Shape and Density.

Table 5
Second set of closely matching attributes from Table 4.

BI-RADS	Age	Shape	Margin	Density	Severity
D	S	Irregular	Circumscribed	Iso	Malignant
A	S	Irregular	Circumscribed	Iso	Benign

Selected Attributes: Age, Shape, Margin and Density.

Step 4: We now go for the second iteration. We consider attributes that are one less than the previous set of attributes. So we now consider three attributes among the four attributes, i.e., from the selected set of closely matching attributes we find the next AAC.

Step 5: We now look up the table to find the corresponding AAC of closely matching tuples. They are shown in Tables 6–9, respectively.

Table 6
AAC I from Table 1.

BI-RAD	Age	Density	Severity
E	S	Iso	Benign
E	S	Iso	Benign
E	S	Iso	Malignant

Class label-Benign Influence factor-1 (age).

Table 7
AAC II from Table 1.

Age	Shape	Margin	Severity
S	Irregular	Circumscribed	Malignant
S	Irregular	Circumscribed	Malignant
S	Irregular	Circumscribed	Benign

Class label-Malignant Influence factor-2 (age, shape).

Table 8
AAC III from Table 1.

Age	Shape	Density	Severity
S	Irregular	Iso	Malignant
S	Irregular	Iso	Benign
S	Irregular	Iso	Malignant
S	Irregular	Iso	Benign

Class label-Undefined Influence factor-2 (age, shape) This table-III is not taken for consideration.

Table 9
AAC IV from Table 1.

Shape	Margin	Density	Severity
Irregular	Circumscribed	Iso	Malignant
Irregular	Circumscribed	Iso	Benign

Class label-Undefined Influence factor-1 (shape) This table-IV is not taken for consideration.

AAC-I
AAC-II
AAC-III
AAC-IV

Table 10
Detailed description of dataset used in the experiment.

Dataset	No. of instances	No. of attributes	No. of classes
Monks'	432	7	2
Iris	150	4	3
All electronics	602	4	2
Wisconsin breast cancer	699	10	2
Hepatitis	155	19	2
Cleveland heart disease	303	13	5
Diabetes	768	20	2
Glass	214	10	7
Chess	3196	36	2
Pima	768	8	2
Soyabean large	307	35	4
Vehicle	946	18	4
Vote	435	16	2
Ionosphere	351	34	2
Telugu vowel	871	3	6
Wine	178	13	3
Yeast	1484	8	10
Zoo	101	17	7

Table 11
A comparison of NB, and NB$^+$ method.

Dataset	NB	Proposed IPMM
Monk 1	74.99	77.80
Monk 2	65.81	68.10
Monk 3	97.23	98.80
Iris	93.13	95.40
All Electronics	91.25	94.10
Wisconsin Breast Cancer	96.4	98.10
Hepatitis	86.3	89.20
Cleveland Heart Disease	82.5	85.12
Diabetes	75.5	77.18
Glass	69.66	73.80
Chess	87.15	92.19
Pima	75.81	82.95
Soybean-Large	91.29	96.10
Vehicle	58.28	66.40
Vote	90.34	93.77
Average	82.37	85.78

Step 6: Attributes that influence the class label are – AGE, SHAPE, and DENSITY (From Step 2.a) Attribute values that influences the class label are:

Age	S, Y
Shape	Irregular
Density	High

Comparing the test data with these values, the attribute values – *S*, Irregular only match. So only 'age' and 'shape' attributes are considered for calculating the Influence Factor in Step 5.

Step 7: Since there are two influencing attributes in the AAC- II we consider the same and assign the class label as **'Malignant'** for the unknown Test sample.

6. Performance analysis

Experiment with the 10-fold cross validation method has been carried out to evaluate the prediction accuracy of proposed NB$^+$, and to compare the experimental results with NB as our benchmarks. Eighteen public datasets have been chosen from the UCI machine learning repository [3]. Information about these datasets is listed in Table 10.

The comparison between NB with NB$^+$ in testing accuracy with the 10-fold cross validation method is listed in Table 11. We have extended the original datasets to contain tuples satisfying the constraint referred to in the problem definition.

It is clear from the experimental results that the average classification accuracy of our proposed IPMM method is better than NB in 10-fold cross validation. The proposed method significantly improves the efficiency of NB by handling the cases where the class probabilities are exactly the same. This is represented in Fig. 1. The graphs below clearly explain that there is a significant rise in the accuracy of the classification. The experimental results show that the average accuracy rate of the proposed method is 85.65%.

7. Summary and conclusion

In this paper an updated procedure has been proposed for the Naïve Bayes algorithm to solve one of its disadvantages. Under the conditions mentioned in this paper equal probability occurs in the class values and so the Naïve Bayes classifier fails to classify the record correctly. When an unknown test sample is to be classified from a finite training dataset with constraints, a better result has been shown to be obtained on the proposed NB$^+$ algorithm is applied. As the method of closely matching tuples is used in the

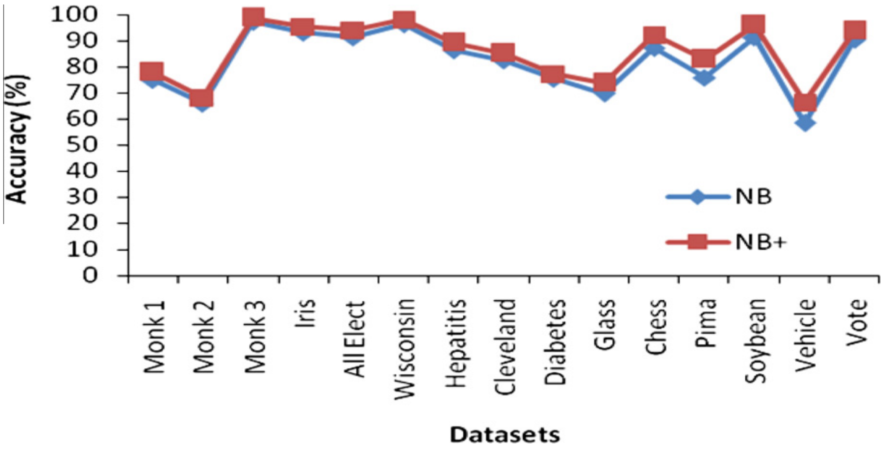

Fig. 1. Predictive accuracy between NB and the Proposed NB$^+$.

NB$^+$ algorithm, a minimal set of attributes (at least one) is necessarily present when the test sample is matched up with the training data. The classification is done considering the influence of the minimal set of attributes on the class label value of the training set. A decision made by the NB$^+$ algorithm based on the above criteria has a higher degree of accuracy than the traditional NB algorithm which makes a random pick of the class label. For achieving a higher degree of accuracy even when the 'minimal set' contains only one attribute, it is necessary to extend the proposed work and construct a more robust and scalable classifier in the near future.

References

[1] S. Appavu alias Balamurugan, Ramasamy Rajaram, Knowledge-Based Sytem for text classification using ID6NB algorithm, J. Knowl.-Based Syst. 22 (1) (2009) 1–7.

[2] M. Bressan, J. Vitria, Improving Naïve Bayes classification using class conditional ICA, Lect. Notes Artif. Intell. 2527 (2002) 1–10.

[3] C. L. Blake,C. J. Merz, UCI Repository of Machine Learning Databases, 2010, available at: <http://www.ics.uci.edu/~mlearn/MLRepository.html>.

[4] A. Denton, W. Perizo, A kernel-based semi-Naïve Bayesian classifier using PTrees. in: Proceedings of the SIAM International Conference on Data Mining, 2004.

[5] L. Diao, K. Hu, Y. Lu, C. Shi, A method to boost Naïve Bayesian classifiers. in: Lect. Notes Comput. Sci., in: Proceedings of the Sixth Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2002, pp. 115–122.

[6] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier underzero-oneloss, Mach. Learn. 29 (1997) 103–130.

[7] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification and Scene Analysis, second ed., Wiley Interscience, NewYork, 2001.

[8] N. Freidman, D. Geiger, M. Goldschmidt, Bayesian network classifiers, Mach. Learn. 29 (2) (1997) 31–163.

[9] Y. Freund, Boosting a weak learning algorithm by majority, Inf. Comput. 121 (2) (1995) 56–285.

[10] Y. Freund, RE. Schapire, A decision theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139.

[11] J. Gama, Iterative Bayes, Theor. Comput. Sci. 292 (2003) 417–430.

[12] H. Huang, C. Hsu, Bayesian classification for data from the same unknown class, IEEE Trans. Syst. Man Cybern. B Cybern. 32 (2002) 37–145.

[13] J.E. Hunter, F.L. Schmidt, Methods of Meta-Analysis, Sage publications, NewburyPark, 1990.

[14] Jingnian Chen, Houkuan Huang, Fengzhan Tian, Shengfeng Qin, A Selective Bayes Classifier for classifying incomplete data based on gain ratio, J. Knowl.-Based Syst. 21 (7) (2008) 530–534.

[15] Jin Xiao, Changzheng He, Xiaoyi Jiang, Structure identification of Bayesian classifiers based on GMDH, J. Knowl.-Based Syst. 22 (6) (2009) 461–470.

[16] E. Keogh, M. Pazzani, Learning augmented Bayesian classifiers. A comparison of distribution-based and classification-based approaches, in: Proceedings of the International Workshop on Artificial Intelligence and Statistics, 1999, pp. 225–230.

[17] A. Kleiner, B. Sharp, A new algorithm for learning Bayesian classifiers, in: Proceedings of the Third IASTED International Conference on Artificial Intelligence and Soft Computing, 2000, pp. 191–197.

[18] R. Kohavi, Scaling up the accuracy of Naïve-Bayes classifiers: a decision-tree hybrid, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 202–207.

[19] T. Kohonen, Self-Organization and Associative Memory, Springer, Berlin, 1989.

[20] I. Kononenko, Semi-Naïve Bayesian classifier, in: Proceedings of Sixth European Working Session on Learning, 1991, pp. 206–219.

[21] P. Langley, S. Sage, Induction of selective Bayesian classifiers, in: Proceedings of the 10th Conference on UAI, 1994, pp. 399–406.

[22] D. Lewis, Naïve Bayes (at forty): the independence assumption in information retrival, in: Proceedings of the 10th European Conference on Machine Learning, 1998, pp. 4–15.

[23] M. Lipsey, D.B. Wilson, Practical Meta-Analysis (Applied Social Research Methods), Sage Publications, London, 2001.

[24] S. Ma, H. Shi, Tree augmented Naïve Bayes ensembles, in: Proceedings of the Third International Conference on Machine Learning Cybernetics, 2004, pp. 1497–1502.

[25] B. Manly, Multivariate Statistical Methods: A Primer, Chapman and Hall, London, 1986.

[26] Michael G. Madden, On the classification performance of TAN and general Bayesian networks, J. Knowl.-Based Syst. 22 (7) (2009) 489–495.

[27] A. Nurnberger, C. Borgelt, A. Klose, Improving Naïve Bayes classifiers using neuro-fuzzy learning, in: Proceedings of the Sixth International Conference on Neural Image Processing, 1999, 154–159.

[28] MJ. Pazzani, Searching for dependencies in Bayesian classifiers, in: Proceedings of Information Statistics and Induction Science, 1996.

[29] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Fransisco, 1988.

[30] PJ. Phillips, EM. Newton, Meta-analysis of face recognition algorithms, in: Proceedings of the Fifth IEEE International Conference on Automatics Face and Gesture Recognition, 2002, pp. 235–241.

[31] C. Ratanamahatana, D. Gunopulos, Scaling up the Naïve Bayesian classifier: using decision trees for feature selection, in: Proceedings of the Workshop on Data Cleaning and Pre-processing, ICDM'02, 2002.

[32] G. Ridgeway, D. Madigan, T. Richardson, J. O'Kane, Interpretable boosted Naïve Bayes classification, in: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 1998, pp. 101–104.

[33] V. Robles, P. Larranaga, E. Menasalvas, M.S. Perez, V. Herves, Improvement of Naïve Bayes collaborative filtering using interval estimation, in: Proceedings of the IEEE WIC International Conference on Web Intelligence, 2003, pp. 168–174.

[34] B. Rosell, L. Hellerstein, Naïve Bayes with higher order attributes, in: Proceedings of the 17th Conference of the Canadian Society for Computational Studies of Intelligence, Lecture Notes in Computer Science, 2004, pp. 105–119.

[35] M. Sahami, Learning limited dependence Bayesian classifiers, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 334–338.

[36] JW. Sammon, A non-linear mapping for data structure analysis, IEEE Trans. Comput. 18 (1969) 401–405.

[37] RE. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227.

[38] RE. Schapire, A brief introduction to boosting, in: Proceedings of the 16th International Joint Conference on Artificial Intelligence, 1999, pp. 1401–1406.

[39] S. Schiffman, M.L. Reynolds, F.W. Young, Introduction to Multidimensional Scaling, Academic, London, 1981.

[40] Shing-Hwa Lu, Ding-An Chiang, Huan-Chao Keh, Hui-Hua Huang, Chinese text classification by the Naïve Bayes Classifier and the associative classifier with multiple confidence threshold values, J. Knowl.-Based Syst. 23 (6) (2010) 598–604.

[41] H.P. Störr, A compact fuzzy extension of the Naïve Bayesian classification algorithm, in: Proceedings of InTech VJFuzzy, 2002, pp. 172–177.

[42] K. Ting, Z. Zheng, Improving the performance of boosting for Naïve Bayesian classification, in: Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1999, pp. 296–305.

[43] A. Tsymbal, P. Cunningham, M. Pechenizkiy, S. Puurone, Search strategies for ensemble feature selection in medical diagnostics, Technical Report, Trinity College Dublin, Ireland, 2003.

[44] A. Tsymbal, S. Puuronen, Ensemble feature selection with the simple Bayesian classification in medical diagnostics, in: Proceedings of the 15th IEEE Symposium on Computer-based-Medical Systems, 2002, pp. 225–230.

[45] R. Vilalta, I. Rish, A decomposition of classes via clustering to explain and improve Naïve Bayes, in: Proceedings of the 14th European Conference on Machine Learning, 2003.

[46] L. Wang, S. Yuan, H. Li, Boosting Naïve Bayes by active learning, in: Proceedings of the Third International Conference on Machine Learning and Cybernetics, 2004, pp. 1383–1386.

[47] Z. Wang, G.I. Webb, Comparison of lazy Bayesian rule and tree-augmented Bayesian learning, in: Proceedings of IEEE International Conference on Data Mining, 2002, pp. 490–497.

[48] G. Webb, J. Boughton, Z. Wang, Not so Naïve Bayes: aggregating one dependence estimators, Mach. Learn. 58 (1) (2005) 5–24.

[49] G. Webb, MJ. Pazzani, Adjusted probability Naïve Bayesian induction, in: Proceedings of the 11th Australian Conference on Artificial Intelligence, 1998, pp. 285–295.

[50] F.M. Wolf, Meta-analysis: Quantitative Methods for Research Synthesis Sage University, Series on Quantitative Applications in the Social Sciences, Sage publications, London, 1986.

[51] H. Zhang, J. Su, Conditional independence trees, Lect. Notes Comput. Sci. 3201 (2004) 513–524.

[52] Z. Zheng, G.I. Webb, Lazy learning of Bayesian rules, Mach. Learn. 41 (1) (2000) 53–84.

[53] Z. Zheng, Naïve Bayesian classifier committees, in: Proceedings of European Conference on Machine Learning, 1998, pp. 196–207.