Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I,
IMECS 2014, March 12 - 14, 2014, Hong Kong

Comparing Performances of Markov Blanket and Tree Augmented Naïve-Bayes on the IRIS Dataset

Haruna Chiroma *Member, IAENG,* Abdulsalam Ya'u Gital, Adamu Abubakar, and Akram Zeki

***Abstract*— This research investigates the performances of the Markov Blanket (MB) and Tree Augmented Naïve-Bayes Network (TAN) of the Bayesian Network structure of the IRIS dataset. For evaluation purposes, the performances of the TAN, and MB classifiers were measured using statistical indices. Experimental results strongly suggested that the TAN is better than MB on training dataset and vise vasa in the test dataset. In the other hand, time computational complexity of both the classifiers was found to be equal. The result obtained in this research is of significance to researchers intending to use Bayesian Network to create a classifier for enhancing the performance of biometrics systems.**

***Index Terms*—Bayesian Network, Markov Blanket, Tree Augmented Naïve-Bayesian Network, IRIS dataset.**

I. INTRODUCTION

BAYESIAN Network can be described as a mathematical object representing a joint probability distribution represented with *J* in which *G* represent a graph annotated with conditional probabilities. A Markov Condition property connected *J* and *G*, a node is said to be conditionally independent of its non-descendants when given its parent. The Markov Blanket of interest *T* (*MB* (*T*)) protected *T* probabilistically from the other attributes which correspond to the neighbor of *T* in the graph of Bayesian Network. $I(X;|Z) \equiv P(T|X,Z) = P(T|Z)$ represent the conditional independence of *X,T* by given *Z*. The Markov Blanket of a variable *T, MB(T),* is minimal for which $I(X;T|MB(T), \forall X \in V - \{T\} - MB(T))$. In Bayesian Network *C* can said to be faithful to joint distribution probability *J* over variable (*V*) such that if and only is $\forall$ dependence entail by the graph of C is as well in *J*. Therefore, the Markov condition creates a close by relationship theoretical probability distribution *J* with the properties of the *G* which correspond to the Bayesian network [1]. Therefore, the Bayesian Network is guided using the *M(B)* algorithm, the *MB(T)* $\forall$ *T* are the initial stage to use for guiding the creation of Bayesian Network [2]. On the other hand, the Tree Augmented Naiv-Bayes Network (TAN) is another Bayesian Network structure currently receiving attention from machine learning and data mining community. The TAN extends the Naive-Bayesian Network in such a way that the attributes to create a tree as shown in Fig 1. The *C* in Fig 1 represents the node of the class and x_1, x_2, x_3, x_4 features forms the tree nodes without their arc. A variation of the Chow-Liu can easily be used for learning such a tree structure of the TAN [3].

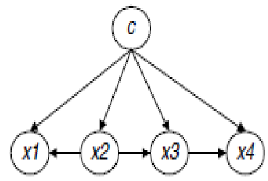

Fig 1. Typical TAN structure

The TAN and MB Bayesian Network have been applying for solving problems in several domains such as classification, image processing, medical diagnostics, prediction among others.

In this paper, we explore the effectiveness of TAN and MB on the UCI IRIS benchmark dataset in order to provide information for researchers on the efficacy of each type of the Bayesian Network structure. The information emanated from the research findings will be of significance to biometric systems developers. Other sections of the paper comprised of Section II provided a brief description of basic theory of Bayesian Network. Section III the IRIS dataset was used to describe. Section IV present step by step experimental description of the study. Section V findings and discussion of the findings are reported. Section VI concludes the paper with the remarks and relevance of the results obtained from the study.

II. THEORETICAL FRAMEWORK

A. *Naiv-Bayesian Network*

In data mining and machine learning community, naiv-Bayesian Network is consider as one of the most efficient inductive learning algorithms and is based on wrapper approach. The Naiv-Bayesian Network is considered a simple classifier with the foundation theory of statistics (Bayes theorem). The Bayesian Network is referred to naïve since it founded based on Bayes Rule, that has a supposition that a feature are conditionally independent from one another [4]. The operations of the Naïve-Bayesian network

H. C. is a PhD research scholar with the University of Malaya, Kuala Lumpur, Malaysia and lecturer with Federal College of Education (Technical), Gombe, Nigeria (corresponding author: +60149576297, e-mail: freedonchi@ yahoo.com).

A. Y. G. is a PhD research scholar with Department of Computer Science, University of Technology Malaysia, Johor Baru, Malaysia and lecturer with Abubakar Tafawa Balewa University, Bauchi, Nigeria (e-mail: asgital@yahoo.com).

A. A. and A. Z. are Assoc. Professor and Professor, respectively with International Islamic University Malaysia, Kuala Lumpur, Malaysia (e-mail: 100adamu@gmail.com).

.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

can be described as follows [4]: Let $X = (x_1, x_2, x_3, ..., x_n)$ be training data points to a particular class label datasetn the training dataset. The probability of each class based on the training data points can be computed using Eqn. (1).

$$P(Y_j|X) = \frac{P(Y_j)P(X|Y_j)}{\sum_{i=1}^{c} P(Y_i)P(X|Y_i)}, j = 1,...,C \quad (1)$$

Where $P(Y_i)$ is the priori probability of class Y_i and $P(Y_j|X)$ represent the class conditional probability density functions. Computation for test cases and prediction from training dataset is computed using Eqn. (2).

$$P(X|Y_j) = \prod_{i=1}^{n} P(X_i|Y_j), j = 1,2,...,C \quad (2)$$

Where X_i is the value of *ith* variable in X and n is the number of variables. Let the number of classes be K and the *ith* class be C_i , the probability distribution over the set of features can be computed using Eqn. (3).

$$P(x) = \prod_{i=1}^{k} P(C_i)P(C_i)P(X|C_i) \quad (3)$$

The efficiency of Naïve-Bayesian Network in classification and learning can be measure, according to the following attributes:

i. Computational efficiency.
ii. Less number of searches due to lower variance.
iii. Tolerance to noise in the dataset.
iv. Effective handling of missing values in the dataset.
v. "Incremental learning because NB functions work from approximation of low-order probabilities that are deduced from the training data. Hence, these can be quickly updated as new training data are obtained".

According to [5] the attribute independence assumption typically require for the performance of Naïve-Bayesian The network makes it unsuitable for attributes that are significantly correlated. Therefore, degrade the performance accuracy of the network if correlated attributes are used. This shows the effect of increasing variances while maintaining a constant bias.

III. The Iris Dataset

For the purpose of our experiments, we collected the IRIS dataset from the well-known UCI machine learning repository [6]. In the IRIS dataset, there are three (3) classes each of fifty (50) instances making a total number of one hundred and fifty (150) rows and five (5) columns comprising of four (4) attributes and one (1) class. No missing values in the dataset. The UCI machine learning repository is considered by the machine learning and data mining community as the benchmark for testing the efficiency and effectiveness of the algorithms.

IV. Experiments

The IRIS dataset is partition into training and test dataset. The data were partition into several ratios in order to ensure consistent findings. Initially the ratio of 50% for training and 50% for testing was used for the experiment. This has eliminated the possibility of bias in the modeling process. In the second partition ratios, the number of observations in the training data is greater than the test dataset as recommended by [7]. In total the dataset was partitioned into four different sizes of training and testing dataset for the conduct of our experimental simulations as can be seen in Table 3. The data were explored to ensure all attributes are in their correct respective columns as expected.

Several simulations were run on the dataset using the TAN and Markov types of Bayesian Network and the best results obtained were recorded and reported in this research. The results that were found to be poor were discarded. The models of the TAN and Markov were analyzed and evaluated. The complete experimental process is depicted in Fig 2 and is implemented using IBM SPSS Modeler Version 15 on a Machine (Nanosec on HP L1750 model, 4Gb RAM, 232.4 GB HDD, 32- bit OS, Intel (R) Core (TM)2 Duo CPU @ 3.00 GHz).

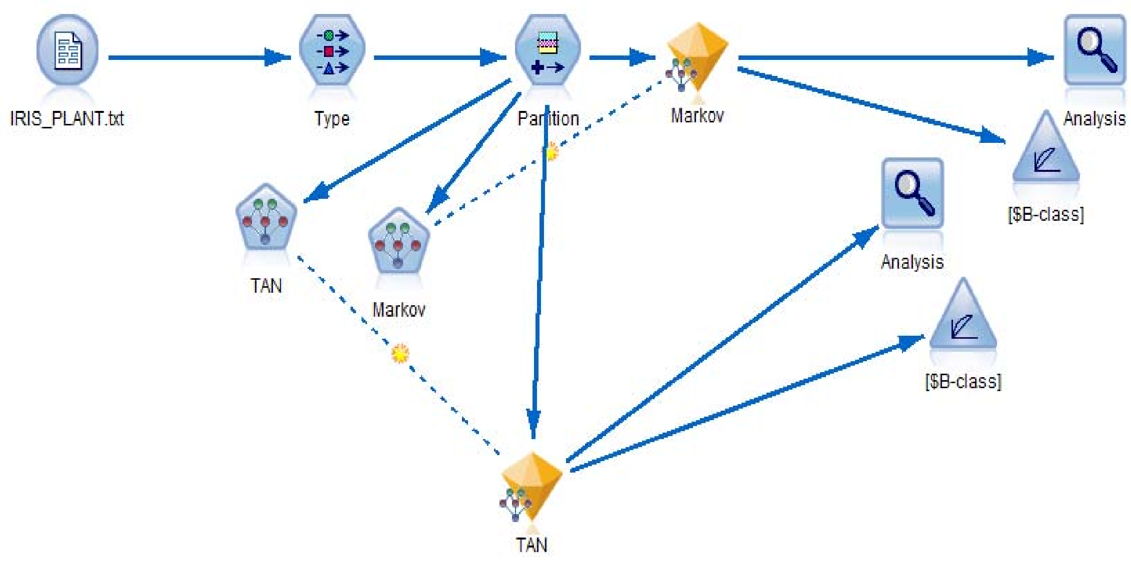

Fig 2. The Propose Framework for comparing performances of MB and TAN structure of Bayesian Network

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. Results and Discussion

The results obtained from the simulations are presented and discussed in this section of the paper for understanding and inferences made from the results. The Bayesian Network for the classification that was performed on the IRIS dataset is presented in Fig 3 showing both the target and predictors. The class of the IRIS determines by four attributes in blue color as displayed in Fig 3.

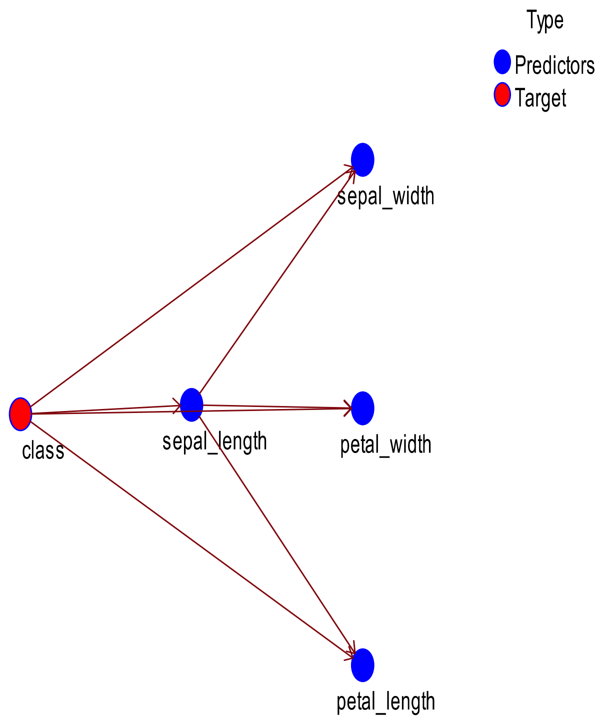

Fig 3. TAN Bayesian Network on IRIS dataset.

The probabilities of the TAN are reported in Table I and that of the MB is presented in Table II showing all the possible combinations derived during the simulations.

TABLE I PROBABILITIES OF THE TAN BAYESIAN NETWORK TYPE OF STRUCTURE

Parents		Probability				
Sepal_length	class	<= 2.48	2.48 ~ 2.96	2.96 ~ 3.44	3.44 ~ 3.92	> 3.92
<= 5.02	Iris-setosa	0.07	0.07	0.60	0.27	0.00
<= 5.02	Iris-versicolor	1.00	0.00	0.00	0.00	0.00
<= 5.02	Iris-virginica	0.00	1.00	0.00	0.00	0.00
5.02 ~ 5.74	Iris-setosa	0.00	0.00	0.25	0.58	0.17
5.02 ~ 5.74	Iris-versicolor	0.00	0.75	0.25	0.00	0.00
5.02 ~ 5.74	Iris-virginica	0.00	1.00	0.00	0.00	0.00
5.74 ~ 6.46	Iris-versicolor	0.00	0.57	0.43	0.00	0.00
5.74 ~ 6.46	Iris-virginica	0.08	0.54	0.38	0.00	0.00
6.46 ~ 7.18	Iris-versicolor	0.00	0.25	0.75	0.00	0.00
6.46 ~ 7.18	Iris-virginica	0.00	0.17	0.83	0.00	0.00
> 7.18	Iris-virginica	0.00	0.40	0.60	0.00	0.00

TABLE II PROBABILITIES OF THE MARKOV BLANKET

Parents	Probability				
class	<= 5.02	5.02 ~ 5.74	5.74 ~ 6.46	6.46 ~ 7.18	> 7.18
Iris-setosa	0.56	0.4	0	0	0
Iris-versicolor	0.05	0.4	0.4	0.2	0
Iris-virginica	0.04	0	0.5	0.2	0.19

The MB Bayesian Network on IRIS dataset are presented in Fig 4.

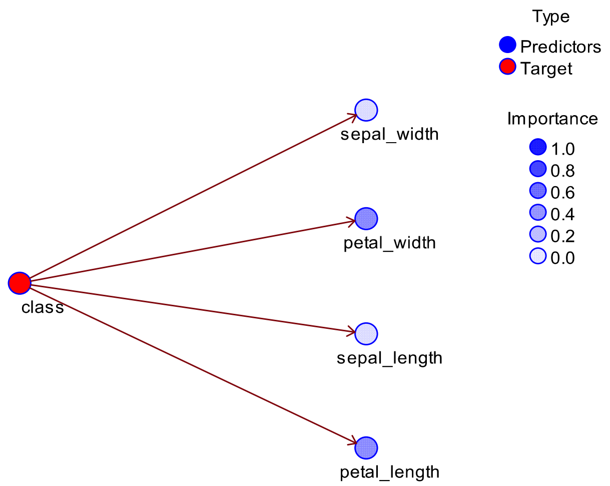

Fig 4. MB Bayesian Network on IRIS dataset

The TAN and MB models were evaluated in order to depict the efficiency of each of the models on the training and test Dataset. Fig 5 shows the performance accuracy of the MB model while Fig 6 shows the performance accuracy of TAN model.

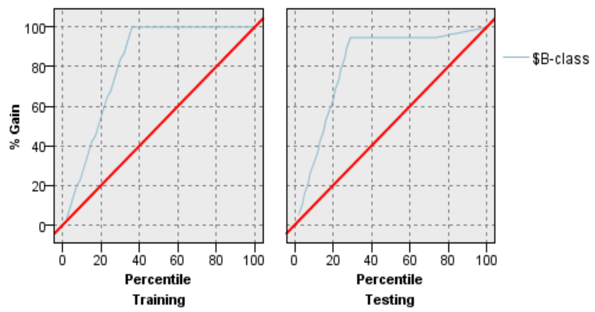

Fig 5. Performance of MB on training-test (60%-40%) dataset.

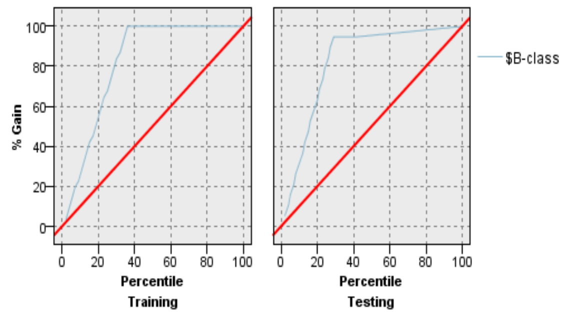

Fig 6. Performance of TAN on training-test (60%-40%)

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

dataset

Fig 5 and 6 were created by observing numerous thresholds for the independence of the attributes included in the datasets. Both the Figures corroborate with the results presented in Table III.

TABLE III COMPARING PERFORMANCES OF MB AND TAN

Data Partition	Algorithm	Correct Test	Wrong Test	Correct Training	Wrong Training
	TAN	77.92	22.08	94.5	5.48
50-50	MB	90.91	9.09	90.41	9.59
	TAN	87.23	12.77	94.17	5.83
70-30	MB	89.36	10.64	91.26	8.74
	TAN	84.21	15.79	94.64	5.36
80-20	MB	86.84	13.16	91.96	8.04
	TAN	79.37	20.63	95.4	4.6
60-40	MB	90.48	9.52	93.1	6.9

The area under the curves shows the accuracy of the classifications. The larger the area the more is the classification accuracy obtained by the model. Both Fig 5 and 6 suggested acceptable level of accuracy. Furthermore, accuracy can be seen in Table III.

The computational time throughout the simulations was zero (0) seconds. For each experiment with the models and data partition ratios were all conducted within zero (0) seconds. Experimental evidence from Table III shows that the classification of MB is better than the TAN on test dataset. In both models, perfect classification which rarely occurs in practice, was not recorded, but the performances of the models were within acceptable range. Cases of misclassification were observed in both models. On the contrary, TAN performs better than the MB on training dataset, though the results in not surprising because it was found in [8] that a model might perform very well on training dataset, whereas its performance can be reduced on the test dataset. The performance of the models is not consistent considering the differences of performance at every stage of the modeling process. In this case a general conclusion cannot be made due to lack of consistence performance for any of the model

.

VI. CONCLUSIONS

In this paper, we compare the performances of TAN and MB of the Bayesian Network structures. The performance metrics used are computation time, classification accuracy in both training and test dataset and receiver operative curve. Experimental evidence suggested that the TAN outperform MB on the training dataset whereas MB performs better than TAN on test dataset. Time computational complexity of both the models was found to be equal. This result could be of significance to researchers intending to use Bayesian Network to create a classifier for used in a biometric system. Different results could be obtained in a different application domain. Therefore, we intend to further this research on multiple datasets from several domains of applications

REFERENCES

[1] I. Tsamardinos, C. F. Aliferis, and A. Statnikov, “Algorithms for Large Scale Markov Blanket Discovery,” American Association for Artificial Intelligence, 2002.

[2] D. Margaritis, and S. Thrun, “Bayesian Network Induction via Local Neighborhoods, Carnegie Mellon University,” Technical Report CMU-CS-99-134, August 1999.

[3] J. Cheng, and R. Greiner, “Learning Bayesian Belief Network Classifiers: Algorithms and System,” Lect. Notes Artif. Intell. vol. 2056, pp. 141-151, 2001.

[4] G. I.Webb, “Na¨ıve bayes,” in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds., pp. 713–714, Springer, New York, NY, USA, 2010.

[5] Z. Zheng, and G.I. Webb, “Lazy learning of Bayesian rules,” Mach. Learn. vol. 41, no. 1, pp. 53-84.

[6] K. Bache, and M. Lichman, “UCI Machine Learning Repository,” [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

[7] I.H. Witten, E. Frank, and A.M. Hall, “Data Mining: Practical Machine Learning Tools and Techniques (3rd Edn.),”. San Mateo: Morgan Kaufmann, 2011.

[8] Y. Jin, “A comprehensive survey of fitness approximation in evolutionary computation,” Soft Comput. Vol.9, pp. 3 – 12, 2005.