Proceedings of National Conference on Machine Learning, 26th March 2019.
ISBN: 978-93-5351-521-8

Implementation of Bayesians Classifier For Probability Estimation of Nutrients Element in The Compost

Savita Mohurle
Research Scholar, Amity University, Mumbai.
Dr. Manoj Devare
Associate Professor, Amity University, Mumbai.
Nakshtra Dhavane
Student, MITACSC, Alandi, Pune.

ABSTRACT

Classifiers are a systematic approach to building a classification model. Classification is the most popular technique in data mining and machine learning. Decision tree, rule-based system, neural networks are some of the classifiers used to determine the uncertainties and probability among the datasets and predicts the desirable results; Likewise, Bayesian network classifiers are statistical classifier that predicts class membership probability. It is based on Baye's theorem, a family of probabilistic classifiers that assist to find the probability and uncertainties in the network. This paper studies the basics of Bayesian classifier model for the analysis and prediction. Also, the paper describes the various applications of a Bayesian classifier network model and illustrates the example for implementing Bayesian classifier to find the likelihood and probability of occurrence of the particular nutrient elements in compost.

Keywords: Bayesian classifier, Bayes theorem, Compost, Conditional probability, Estimation, Probabilistic classifier, Random variable.

INTRODUCTION

To calculate the uncertainties using probability, Bayesian approach is found to be most appropriate. Probability is used in numerous applications in day to day life. It can be considered as an event by chance. A Bayesian network is a directed acyclic graph (DAG) in which the nodes denote random/chance variables, and the arcs or lack of them denote the qualitative relations among the variables (MaCann,2006). First, Bayesian networks used to model the probabilistic influence of a set of variables on another variable in the network. Given the probability of parents, the probability of their children can be calculated. Second, Bayesian networks can cope with the missing data problem (Yuan, Lim and Lu, 2011). Nowadays, a Bayes classifier methodology is used in many applications such as engineering, medical, finance, etc. (Mengersen and Pitchforth, 2016). Currently, the applications of BN has moved from engineering to medical diagnosis, business, finance and management, safety analysis to decision making adaptive testing and many more. This paper illustrates an example of how Bayesian classifier is used to find the likelihood and probability of the composition of elements present in the compost sample.

LITERATURE REVIEW

(Wu, Haugn, Lin, Liao, Peng, Hung, Wu, Tu and Chien, 2013), has explained an application of randomized clinical trials, the trails are reviewed and meta analysis is done using Bayesian network . (Khakad, 2011), proposed MRE method to find partial instance of the target variables to give the best explanation of the evidences and maximize the GBF. (Campos**,** Fernandez Gamez and Puerta, 2002), introduced algorithm based on meta heuristic for optimization of ant colony problem. They tried to tackle all the learning problem

and compare performance of algorithm. (Krishnamachari and Iyengar, 2003), propose a distributed solution for wireless sensor network to detect environmental features by developing an algorithm to detect and correct faults in wireless network. The resultant showed that 85-95% faults are corrected and only 10% remains faulty. (Fenton, Neil, and Lagnado, 2016), has extended the work of the author on object oriented Bayes Network for complex argument structure. (Cai Liu, Fan ,Zhang, Liu, Yu, and Ji, 2014), proposed fault diagnosis system for multiple sources that acquire knowledge and probabilistic representation of the uncertainties. (Pitchforth, and Mengersen , 2012) to extends the work of previous other authors to measure latent variables, establish a confidence even if there is no data available using Bayesians network. (Khakzad, Khan, and Amyotte, 2011) applied BN in safety analysis hence stating that BN is the technique used in safety analysis in many accidental scenario.

MATERIAL

Finding probability distribution is equally helpful when considering its application in medical diagnosis and health, financial management and banking, fraud detection and forgery. In fact it has already found their application in health outcomes, decision and analysis and probability distribution of random events in vide area (Arcienega, 2009). Classification is the task to learn a concept that maps attribute x(value in dataset) with y(desirable value) (Cai Liu, Fan ,Zhang, Liu, Yu, and Ji, 2014). A Bayesian network is a representation of a joint probability distribution of a set of random variables with a possible mutual causal relationship. The network consists of nodes and a conditional probability distribution in each of the node. Bayesian classifiers are also known as Naive Bayesian classifier that states that the values of attributes are independent to each other. The naive Bayesian classifier works as follows:

1. Divide the attributes into some range, to classify the new tuple.
2. Estimate the priori probability $P(C_j)$ for each class with available training data(occurrence of each class in training data)
3. Estimate the posterior probability $P(C_j|x_i)$. Probability that xi belongs to C_j.
4. Estimate the probability of occurrence of each attribute $P(x_i)$.
5. Use the values in step 2,3 and 4 to find the highest probability of the new tuple.

Illustrating Example

Consider compost sample data in table 1. Data consists of 10 compost sample having seven mineral nutrients available. Also, Resultant value1 and Resultant value 2estimated based on composition of each element in sample.

Consider resultant value1 as classification results. There are four tuples having resultant value1as Good, three tuple having resultant value1as Excellent and Moderate respectively. Since, Naïve Bayesian classifier is a type of supervised learning, the output resultant value1 is decided on the standard requirement specifications that any potential compost should have. According to the standard requirement from (Fertilizers Order, 1985), a compost sample should contain <5.0, 1000, 100, 10.00, 1.0, 2.0, 1.40> as resultant for <Cd, Zn, Pb, Ar, N, Ph, K> respectively. Following illustrate the example of finding the actual value of cadmium element. The Bayesian classifier is used to assess the likelihood of availability of each attribute in compost sample. To facilitate classification, the value of Cd attribute is divide into seven ranges such as <(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7)>. Table 2 shows the count and subsequent probability of Cd element. Depending on the range the probability associated with Cd, the probability of an element is determined. Naive bayes classification can be very efficiently used as a supervised learning. It estimates parameter based on maximum likelihood and determines the class membership probability, such as the given tuple belongs to a particular class (Ahmet and Olcay, 2014).

Table 1. Compost Sample

ISBN: 978-93-5351-521-8

Sample	Cd	Zn	Pb	Ar	N	Ph	K	Resultant value 1	Resultant value 2
S1	3.55	999	100	9.8	0.99	0.1	1.3	Good	Moderate
S2	4.15	1100	99	8.1	1	0.3	1.99	Excellent	Good
S3	3.15	1018	101.99	10.33	0.5	0.4	2	Good	Moderate
S4	4.99	1199	98.99	9.11	0.76	0.55	1.5	Good	Moderate
S5	5.98	1001	100.78	10.54	0.82	0.34	1.2	Excellent	Good
S6	4.15	998.8	102.3	11.16	1	0.23	0.99	Excellent	Good
S7	3.99	993	98.67	7.11	0.89	0.45	0.98	Good	Moderate
S8	3.99	889.2	101.33	8.99	0.77	0.21	1.5	Moderate	Impotent
S9	6.35	888	100.44	9.99	1.1	0.4	1.2	Moderate	Impotent
S10	5.55	902	93.55	9.1	0.66	0.4	1.3	Moderate	Impotent

Table 2. Count and subsequent probabilities of Cd

Attribute	Value	Count			Probability		
		Good	Excellent	Moderate	Good	Excellent	Moderate
Cd	0,1	0	0	0	0	0	0
	1,2	0	0	0	0	0	0
	2,3	0	0	0	0	0	0
	3,4	3	0	1	¾	0	1/3
	4,5	1	2	0	¾	2/3	0
	5,6	0	1	1	0	1/3	1/3
	6,7	0	0	1	0	0	1/3

Initially, the prior probability is calculated as follows:

P(Good)=4/10=0.4
P(Excellent)=3/10=0.3
P(Moderate)=3/10=0.3

Using values in table2 and the calculated priori probabilities a new tuple for Cd is classified. Suppose new tuple is, t=<s11, 4.8, 999, 99, 8.1, 0.99, 1.5, 1.2>. Estimate the p(t|Good)=1/4=0.25, p(t|Excellent)=1/3=0.3333, p(t|Moderate)=1/3=0.3333. Therefore, Likelihood of getting output index Good=0.25*0.4=0.1, Excellent index=0.3333*0.3=0.0999, Moderate index=0.333*0.3=0.0999.
Summation of p(t)=0.1+0.0999+0.0999=0.2998
Actual probability of p(Good|t)=(0.25*0.4)/0.2998=0.333, p(Excellent)=(0.333*0.3)/0.2998=0.333 and p(Moderate|t)=(0.333*0.3)/0.2998=0.333.

RESULTS

The probability value shows that proportion 4.8 for Cadmium (Cd) in sample s11 is accurate for any compost sample having same proportion of Cd, since the resultant value comes to be same for Good, Moderate as well as Excellent for subsequent probability. That is for any compost sample having 4.8 proportion considered to be near desirable value 5.0 as well as to permissible value. Naive Bayes approach

is very straight forward to use. Since data set is divided into range it becomes effortless to determine the probabilities. In the sense 4.8 as an attribute value for Cadmium mineral nutrient is considered in anyways. Since the probability for resultant value Good, Moderate and Excellent comes out to be 0.33. Finally, the proportion of Cd is 4.8 is classified to be the excellent proportion having highest probability. Likewise, the count value and probability value of each and every mineral nutrient in whole sample of compost is calculated and thus, the actual probability of each and every element present in compost is determined.

CONCLUSION

The Naive Baye's classifier benefits to find the probability and uncertainties in the data in vide range of applications. Even though there are certain other concepts available to find the uncertainty or probability of occurrence, Baye's classifier responds very well in finding out the relationship between casual random variables. A humble attempt is also made to explain the concepts of Baye's classifier by illustrating the example of compost.

REFERENCES

Arcienega M., A. A. (2009). A Bayesian network approach combining pitch and spectral envelope features to reduce channel mismatch in speaker verification and forensic speaker recognition. *Interspeech 2005*, 4-8.

Cai B., L. Y. (2014). Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network. *Applied Energy, Elsevier*, 1-9.

Department of Agriculture and Cooperation. (1985). The Fertilizers (Control) Order 1985.Retrieved from http://krishi.bih.nic.in/Acts- Rules/Fert_Order_1985.pdf

Fenton N., N. M. (2016). A General Structure for Legal Arguments About Evidence Using Bayesian Networks. *Cognitive Science, A Multidispalinary Journal*, 61-102.

Mengersen, K. and Pitchforth, J.(2016). A proposed validation framework for expert elicited Bayesian Network. *Expert Systems, Elsevier*, 1-27.

Khakzad N., K. F. (2011). Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering and System Safety,Elsevier*, 925-932.

Krishnamachari B., and Iyengar S., "Distributed Bayesian Algorithm for fault-Tolerant Event Region Detection in Wireless Sensors Network",

Luis M. de Campos, J. M.-L. (2002). Ant colony optimization for learning Bayesian Network. *International Journal of Approximate Reasoning, Elsevier*, 291-311.

McCann R.K., M. B. (2006). Bayesian belief networks: applications in ecology and natural resource management. *Cananda Journal for Resource*, 3053-3062.

Okutan A., Taner O.Y., "Software Defect Prediction Using Bayesian Networks", Empirical Software Engineering, Vol. 19, No.1,, 154-181.

Tibhuvan A.P., Tibhuvan P.P., and Gade J.G.,(2015), "Applying Naive Bayesian Classifier for Predicting Performance of a student using WEKA", *Advances in Computational Research*, Vol. 7, No.1, 2015.

Wu H. Y., H. J. (2013). Comparative effectiveness of renin-angiotensin system blockers and other antihypertensive drugs in patients with diabetes: systematic review and bayesian network meta-analysis. *BMJ*, 2-15.

Yuan C., L. H. (2011). Most Relevant Explanation in Bayesian Networks. *Journal of Artificial Intelligence Research* , 309-352.