International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 89-93, 2012

Development of Maize Expert System Using Ada-Boost Algorithm and Naïve Bayesian Classifier

M.S. Prasad Babu
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

Venkatesh Achanta
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

N.V.Ramana Murty
Dept.of CS
Rayalaseema University,
Kurnool,
A.P,India

Swapna.K
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

Abstract: Machine learning Recent works on ensemble [1] methods like Adaptive Boosting have been applied successfully in many problems. Ada-Boost algorithm running on a given weak learner several times on slightly altered data and combining the hypotheses in order to achieve higher accuracy than the weak learner. This paper deals with the design and development of an expert system to advice the farmers in villages through online. An expert system is a computer program, with a set of rules encapsulating knowledge about a particular problem domain. This is a web based application developed using machine learning techniques. In the present paper, Ada-Boost algorithm technique has been considered and applied to generate conclusion based on the given training data provided by human expert. Here a rule based expert system and machine learning systems are integrated to form the proposed Ada-Boost Algorithm Using Naïve Bayesian Classifier on Maize expert Advisory System. The proposed system examines the symptoms provided by the user and process the information based on the training data and determines the diseases effected to the Maize crop. The system shows a global solution for recognizing the diseases in Maize crop cultivation and also suggests the corresponding treatments to the diseases. This expert system is a web based application for online users with java as front end and MYSQL as backend.

Keywords: Expert Systems, Rule-Based System, Machine Learning, Ada-Boost, Naïve Bayesian Classifier, Maize, JSP and MYSQL

1. INTRODUCTION

A. Maize Crop Information

Maize known in many English-speaking countries as corn or mielie/mealie is a grain domesticated by indigenous peoples in Mesoamerica in prehistoric times. The leafy stalk produces ears which contain seeds called kernels. Maize kernels are technically a fruit but are used in cooking as a vegetable or starch. The Olmec and Mayans cultivated it in numerous varieties throughout central and southern Mexico, cooked, ground or processed through nixtamalization. Between 1700 and 1250 B.C, the crop spread through much of the Americas. Sugar-rich varieties called sweet corn are usually grown for fresh consumption while field-corn varieties are used for animal feed and as chemical feed stocks.

B. Expert Systems

An expert system [5] is software that uses a knowledge base of human expertise for problem solving, or to clarify uncertainties where normally one or more human experts would need to be consulted. Expert systems are most common in a specific problem domain, and are a traditional application and/or subfield of artificial intelligence (AI). A wide variety of methods can be used to simulate the performance of the expert; however, common to most or all are: 1) the creation of a knowledge base which uses some knowledge representation structure to capture the knowledge of the Subject Matter Expert (SME); 2) a process of gathering that knowledge from the SME and codifying it according to the structure, which is called knowledge engineering; and 3) once the system is developed, it is placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or as a supplement to some information system. Expert systems may or may not have learning components. A series of Expert advisory systems [12], [13], [15] were developed in the field of agriculture and implemented in www.indiakisan.net[14].

C. Machine Learning

Machine learning[2, 3, 4 and 6], a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. It is a very young scientific discipline used in various areas including Robotics, Machine Vision, etc. The First Machine Learning Workshop was taken place in 1980 at Carnie-Mellon University (USA).The goal of machine learning is to program computers to use training data or past experience to solve a given problem. Effective algorithms have been invented for certain types of learning tasks. Many practical computer programs have been developed to exhibit useful types of learning and significant commercial applications have begun to appear. Machine learning refers to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. Some of the machines learning algorithms are Genetic Algorithm, Decision Tree Algorithm, Optimization Algorithm, Adaptive Boosting Algorithm, Bagging Algorithm and Particle Swarm Optimization Algorithm, Bayesian Classifier Algorithm, ID3 and C4.5 Algorithm.

D. Adaptive Boosting (Ada-Boost) Algorithm

Ada-Boost, [7, 8, 11] short for Adaptive Boosting, is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire. It is a meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance. Ada-Boost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. Ada-Boost is

sensitive to noisy data and outliers. However in some problems it can be less susceptible to the overfitting problem than most learning algorithms. Ada-Boost calls a weak classifier repeatedly in a series of rounds t=1,2,…..T from a total T classifiers. For each call a distribution of weights Dt is updated that indicates the importance of examples in the data set for the classification. On each round, the weights of incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

The pseudo code for Ada-Boost algorithm is given as below

• **Input**: a set S, of 'm' labeled examples: S= ((xi,yi), i=(1,2,…,m)), with labels in Y.

• **Learn** (a learning algorithm)

• **A constant L.**

[1] Initialize for all i: wj(i)=1/m // initialize the weights

[2] for j=1 to L do

[3] for all i: // compute normalized weights

$$p_j(i) = \frac{w(i)}{\sum_i^m w(i)}$$

[4] hj: =Naïve-Bayesian(S,pj) // call weak Learn with normalized weights

[5] Calculate the error of hj

$$\varepsilon_j = \sum_i p_j(i)[h_j(x_i \neq y_i)]$$

[6] if $\varepsilon_j > \frac{1}{2}$ then

[7] L=j-1

[8] go to 12

[9]

$$\beta_j = \frac{\varepsilon_j}{1-\varepsilon_j}$$

[10] for all i: // compute new weights

$$w_{j+1}(i) = w_j(i)\beta_j^{1-[h_j(x_i-y_i)]}$$

[11] end for

12] Output:

$$h_{final}(x) = \arg\max_{y \in Y} \sum_{j=1}^{L} \left(\log \frac{1}{\beta_h}\right)[h_j(x=y)]$$

E. Naïve Bayes Classifier (Weak Classifier)

Naïve Bayes Classifier is a simple probabilistic classifier based on Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". Depending on the precise nature of the probability model, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naive Bayes models uses the method of maximum likelihood; in other words, one can work with the Naive Bayes model without believing in Bayesian probability or using any Bayesian methods. In spite of their naive design and apparently over-simplified assumptions, Naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, analysis of the Bayesian classification problem has shown that there are some theoretical reasons for the apparently unreasonable efficacy of Naive Bayes classifiers [9]. Still, a comprehensive comparison with other classification methods in 2006 showed that Bayes classification is outperformed by more current approaches, such as boosted trees or random forests [10].

2. PROPOSED MAIZE EXPERT ADVISORY SYSTEM

A. Maize Crop Information

The Proposed Ada-Boost Algorithm uses the Naïve-Bayes classifier as weak learner and it uses the training data and the weights are initialized based on the number of classifiers i.e., the weights of the each class is equal to the fraction of the total number of classifiers. Select 'T', the number of rounds the algorithm has to run iteratively by adjusting the weights. In each round the weak learner is called based on the given input and the weights for each classifier and it generates a new hypothesis 'hj' in each hypothesis and the weight and the error is calculated based on the obtained hypothesis and based on the error value obtained the new weights are calculated by using the formula given below

$$w_{j+1}(i) = w_j(i)\beta_j^{1-[h_j(x_i-y_i)]}$$

where βj is error coefficient.

The weak learner is called by using the new weights. The process is repeated until the error value greater than ½ or the number of iterations completes. And finally, the hypothesis value is calculated by using the given formula.

$$h_{final}(x) = \arg\max_{y \in Y} \sum_{j=1}^{L} \left(\log \frac{1}{\beta_h}\right)[h_j(x=y)]$$

The flow diagram of the proposed Ada-Boost algorithm used in development of this Expert Advisory System is shown in fig. 1.

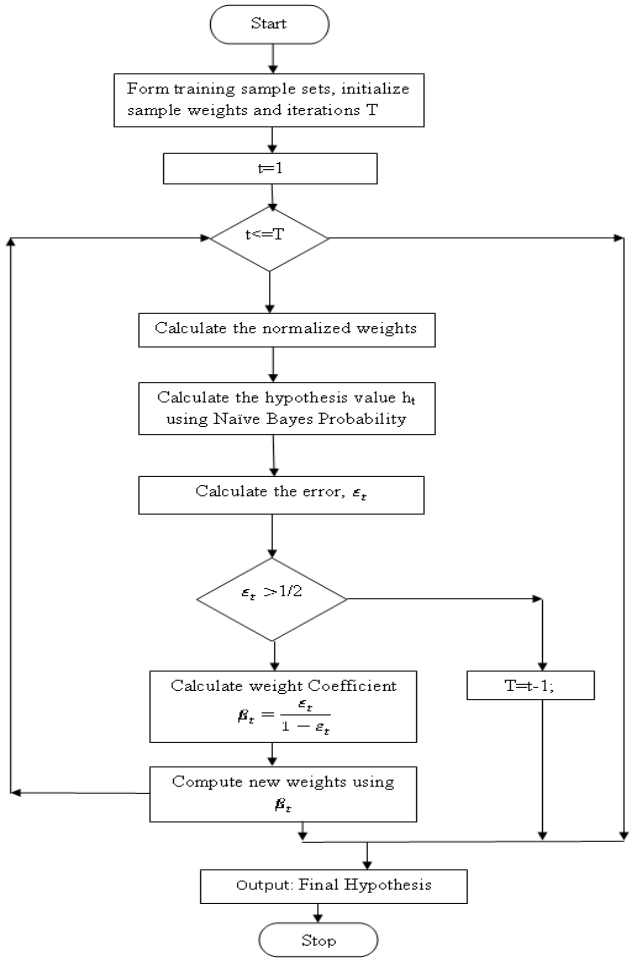

Figure 1 Flow chart of the Ada-Boost Algorithm

B. Simple Example

The working of the proposed system is explained by considering the 10 symptoms as input. It is explained as follows

- Encode Solution: Just use 10 bits (1 or 0).
- Generate input.

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	0	1	0	0	0	1	0	1	0

- Initialize the weights wi based on the classifiers. Consider there are 5 classifiers, wi= 1/5.
- Select the value for 'T', the number of iterations.
- In each and every, iterations the hypothesis value 'hj' is to be calculated.

The probability densities for each disease is calculated using the naïve Bayesian classifier as follows

P(Disease1/s1,..s10)= P(Disease1)*P(s1/Disease1)*P(s2/Disease2)….P(s10/Disease)

By using the above equation for the given input string

P(Corn Streak Virus/1,0,1,0,0,0,1,0,1,0)= 0.20000002

P(Sorghum Down Mildew/1,0,1,0,0,0,1,0,1,0)=0.0

P(Postfloweringstalk rot/1,0,1,0,0,0,1,0,1,0)=0.029626261

P(Phaesophearia Leaf spot/1,0,1,0,0,0,1,0,1,0)=0.12922

P(Alternaria Leaf Spot/1,0,1,0,0,0,1,0,1,0)=0.122

The Probability value for disease Corn Streak Virus is greater than all the remaining diseases hence the hypothesis generated by disease name as "Corn Streak Virus".

- The error value is calculated by adding the probabilities value of the remaining diseases with their corresponding weights.
- Based on the error value the algorithm is repeated repeatedly for 'T' times by adjusting the weights.

The hypothesis values and error values of the given input for **Corn Streak Virus** using Ada-Boost algorithm are shown as below.

Value of 'T'	Hypothesis Value	Error Value
1	0.20001	0.455
2	0.22001	0.346
3	0.22286	0.320
4	0.2272	0.312

Figure 2 Computations Performed During Ada-Boost Algorithm

The final hypothesis generates the disease having the hypothesis value is greater than all other diseases.

Thus for the given example the "Corn Streak Virus" disease is generated.

3. DATABASE DESIGN

This section explains the database design used for the development of the expert system. It explains about the different rules stored in the knowledge base. Generally the rules of the form,

Rule1: S1=0, S2=1, S3=0, S4=1, S5=1, S6=0, S7=1, S8=0, S9= 1, S10=0 resultant disease may be "CORN STREAK VIRUS".

Rule2: S1=1, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1 resultant disease may be "PRE FLOWERING STALK ROT".

Rule3: S1=0, S2=1, S3=1, S4=0, S5=1, S6=1, S7=0, S8=1, S9=1, S10=1 resultant disease may be "SORGHUM DOWNY MILDEW".

Rule4: S1=0, S2=1, S3=1, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be "POST FLOWERING STALK ROT".

Rule5: S1=0, S2=0, S3=0, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be "ALTENARIA LEAF SPOT".

Rule6: S1=1, S2=1, S3=1, S4=1, S5=0, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be "TURICUM LEAF BLIGHT".

Rule7: S1=0, S2=0, S3=1, S4=0, S5=1, S6=1, S7=1, S8=0, S9=1, S10=1 resultant disease may be "FLEA BEETLES AND FLEA ROOTWORMS".

Rule8: S1=0, S2=0, S3=0, S4=1, S5=0, S6=1, S7=1, S8=1, S9=1,S10=0 resultant disease may be "PHYLLOSTICTAL LEAF SPOT".

Rule9: S1=1, S2=1, S3=0, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be "CHARCOAL STALK ROT ".

Rule10: S1=0, S2=1, S3=0, S4=1, S5=0, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be "MAIZE FINE STRIPE VIRUS".

4. RESULTS

Description: In this screen shot, the user can submit the observed symptoms to the maize advisory system through online by selecting the appropriate radio buttons for the processing of the symptoms observed.

From the above screenshot, the following result is observed

Effected With: Sorghum Downy Mildew

Cure is: Spray carbendazim 1.5g and use metalaxyl MXL

Figure 2 Screen for selecting symptoms in Maize Expert System

Figure 3 Displaying advices to the farmer

5. CONCLUSION

An Expert Advisory System entitled "Maize Expert Advisory System Using Ada-Boost Algorithm" is developed using Java Server Pages (JSP) and MYSQL database as backend. This system generates advices based on the symptoms given by the farmer and gives appropriate suggestions to improve the productivity of the crop. This algorithm enhances the performance of the weak learner in iterations by adjusting the weights and reducing the misclassification error values. Thus the performance of the system is enhanced.

6. REFERENCES

[1] Dietterich, T. G., 2000, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Machine Learning*, pp *139-158.*

[2] . Reviews of "Machine Learning by Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell", Tiago Publishing Company, 1983, ISBN 0-935382-05-4.

[3] Corinna Cortes and Vladimir Vapnik. "Support-vector networks", Machine Learning, 20(3):273–297, 1995.

[4] C.A.Coello, Gary B.Lamont and David A.Van Veldhuizen, "Evolutionary Algorithms for Solving Multi-Objective Problems", *2nd Edition, Springer*, 2007.

[5] Dr. B D C N Prasad, P E S N Krishna Prasad and Y Sagar, "A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)" *–CCSIT 2011, Part I, CCIS 131*, pp 570– 576.

[6] Rennie J, Shih L, Teevan J, and Karger D, "Tackling The Poor Assumptions of Naive Bayes Classifiers," In Proceedings *of the Twentieth International Conference on Machine Learning (ICML).* 2003.

[7] Yoav Freund and Robert E. Schapire, "Experiments with a new boosting algorithm", In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 148–156, 1996.

[8] A. J. M. Abu Afza, Dewan Md. Farid, and Chowdhury Mofizur Rahman, "A Hybrid Classifier using Boosting, Clustering, and Naïve Bayesian Classifier", *World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 3,*105-109, 2011.

[9] Rish Irina, "An empirical study of the naive Bayes classifier", *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.*

[10] Atsushi Takemura, Akinobu Shimizu, and Kazuhiko Hamamoto, "Discrimination of Breast Tumors in Ultrasonic Images Using an Ensemble Classifier Based on the AdaBoost Algorithm With Feature Selection", *IEEE Transactions on Medical Imaging, vol. 29, no. 3,* 2010.,

[11] Atsushi Takemura, Akinobu Shimizu, and Kazuhiko Hamamoto, "Discrimination of Breast Tumors in Ultrasonic Images Using an Ensemble Classifier Based on the AdaBoost Algorithm With Feature Selection", *IEEE Transactions on Medical Imaging, vol. 29, no. 3,* 2010.,

[12] Prof. M.S. Prasad Babu, N.V. Ramana Murty, S.V.N.L.Narayana, "A Web Based Tomato Crop Expert Information System Based on Artificial Intelligence and Machine learning algorithms", *International Journal of Computer Science and Information Technologies, Vol. 1 (1), (ISSN: 0975-9646).*, 2010, pp6-15.

[13] Prof. M.S. Prasad Babu, Mrs. J. Anitha, K. Hari Krishna, "A Web Based Sweet Orange Crop Expert System using Rule Based System and Artificial Bee Colony Optimization Algorithm" , *International Journal of Engineering Science and Technology ,vol.2(6)*,2010.

[14] www.indiakisan.net

[15] Prof. M.S. Prasad Babu, N.Thirupathi Rao, "Implementation of Parallel Optimized ABC Algorithm with SMA Technique for Garlic Expert Advisory System", *International Journal of Computer Science, Engineering and Technology (IJCSET), Volume 1, Issue 3,* October 2010.