Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

Accepted January 2014

Research Article

Improving the Methods of Email Classification through the Fuzzy Decision Tree

Enayat Bayati, Mehdi Sadeghzadeh, Farshad Kumarci

Department of Computer Engineering, Science and Research Branch of Esfahan, Islamic Azad University, Esfahan, Iran

Abstract

The Internet has dramatically changed the relationship among people and their relationships with others. Email is the service, providing by the Internet today for its own users; this service has attracted most of the users' attention due to the low cost. Along with the numerous benefits of Email, one of the weaknesses of this service is the continuous enhanced of the received emails. The rapid expansion of this service among the Internet users has caused that some of the individuals to exploit it resulting in the spread of spam. In this paper, we introduce a new method to detect and classify the spam. We increased the precision of Email classification through FID3 decision tree and compared the results with two methods, SVM and Naïve Bayesian, by F-Measure and precision criteria; and finally succeed to make an acceptable balance between the spam detection error instead of valid email and vice versa.

Keywords: Data Mining, Email classification, spam, decision tree, fuzzy decision tree,ID3

I. Introduction

E-mail spam has become a serious problem today, so that spam contributes in 75-80% of e-mail volume [MAAWG 2006]. Spam creates several problems some of which directly lead to the economic losses. Specifically, spam causes the traffic and destroys the storage space and computing power. As the result of spam, the users spend more time for selecting and removing the spam Emails and thus leads to the emotional abuse and lack of security in them; eventually the spam causes the legal problems such as advertising, pornography, pyramid and scam projects such as phishing. Ferris Research Institute has estimated that the economic loss, resulting from unsolicited Email (spam), is about $ 50 million [MAAWG 2006].

Spam entered in the Internet in 1990 when it became a disaster which was expanding every hour. The users spend more time to read and sort these types of email and it is not economical; Furthermore, the spam may occupy a lot of space on the server causing a lot of problems for most of the web pages with thousand users. The technical issue is another issue associated with the spam. Most of the type of spam can be dangerous and may include the viruses, Trojan horses or dangerous software leading to the failures in computers and networks. Therefore, the programs are needed to automatically filter these troublesome Emails; in fact the spam filter is a computer program with the ability to classify the Emials.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

Nowadays, most of the activities for spam filtering are based on the techniques such as Naïve Bayes classification, neural networks, etc. In this paper, we introduce a new approach to classify the emails based on j48 decision tree and FID3fuzzy decision tree. We have utilized the spambase[spambasedadaset] and Weka software for building the decision tree.

The following sections are presented: Section 2 represents the research background and relevant studies; section 3 presents the article idea through the decision tree and fuzzy decision tree; Section 4 provides the tests through the proposed framework, and finally Section 5 presents the conclusion and future works.

II. Research background and related works

Numerous studies have been conducted in the field of spam and they can be classified as the details display, details selection and classification.

A. Details display and selection

Text display is an important option in the application of text classification. The text is usually obtained from the body of message in filtering the spam although the topic or even the header of Email can also be taken into account. Bag of words or vector space is one of the most common methods of display. Another method is to apply the character n-gram model in which the classified characters are obtained by scrolling the window with the certain size *n* on the text. Another way is to utilize the word n-gram and the third method of display is *tf-idf* method by which the value of $x_i$is obtained for t_i based on the equation 1 in which $n_{t_i,d}$ is the number of occurred t_i in the document *d*,n_{t_i} is the number of document in which there is t_iIs and D_{tr} is the set of document.

$$x_i = n_{t_i}\log(\frac{|D_{tr}|}{n_{t_i}}) \qquad (1)$$

In equation [Goodman 2004], the words are extracted by *n*-gram method and then some of them are classified and displayed by the name of group. *PC*-KIMMO system is applied in order to convert the word to its root. The authors of this article utilized *CGP* Model for building the class and also *ESP* Model for classification of Emails. The author in [Watson 2004] utilized the character *n-gram* for extracting the details and displayed it as the binary system. An online system based on *SVM* Algorithm is designed in this paper and applied on the data; and then the obtained results are investigated and compared with the results of bag of words model. The words are applied as the details for email classification in [Çıltık et al 2008]. This article investigates the effect of word size on the precision of classification. The details selection methods, *DF*, *IG* and ratio of inconsistencies are also utilized in [Chang et al 2009].

B. Classification

Different combination of Email, Email header, Email subject and body are applied in [Wang et al 2006] for concentration and extraction of details. Its author has taught the system with three learning algorithms, SVM, NB and K-NN, after extraction of details Three and then has investigated and compared the results obtained from those three algorithms. Finally, a method combined of TF-IDF and SVM algorithms is utilized; the former is for extracting the details from the Email and the latter is for predicting whether the Email is spam or not. Four

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

algorithms, NB machine learning, neural network, SVM and vector machine, are utilized in [Hu et al 2010] for classification of spam, and then obtained results are investigated. [Blachnik et al 2009] classified the data base into six subsets in order to classify the spam and then each of the subsets is divided into the test and educational groups. K-NN is utilized for model education. The final result is selected based on the highest number at the final stages.

A combined system from the decision tree, SVM, and back-propagation neural network is applied in [Lai 2007] based on machine learning. The features of system consist of 14 cases given to this system after the preprocessing stages, and the method with highest vote is utilized for combining the results of three algorithms in order to find the final result indicating the spam or normal Email. The precision of implementing the system is reported equal to 91.78% which indicates the higher precision compared to the implementation of each three algorithms.

III. Spam filtering through J48 decision tree and fuzzy decision tree

A. General Approach

The first stage covers building a smart decision tree; j48 tree is utilized in this regard and we identify whether the Email is spam or not; the second stage represents the tree fuzzification and we reinvestigation of result. First, we collect an appropriate dataset based on which the decision tree is built. This dataset should include the characteristics of spam and valid Emails. We applied the spam base [spam base dataset]. We evaluated a number of decision trees in Weka and decided to utilize j48tree. j48 tree is an implementation of C4.5 decision tree. It accepts the input with the format Arff. Spam base includes 4601 emails consisting of 1813 spam (39.4%) and 2788 valid emails (60.6%). We applied 4101 ones for training and 500 emails for testing.

B. Architecture FSF (Fuzzy Spam Filtering)

Figure 1 shows the general framework of this research for spam filtering called as the FSF. we covert the training data into the format Arff and build a decision tree based on the training data,and finally the test Emails are given to it and it is investigated whether it is spam or valid Email. All leaves have the values equal to 0 or 1 in the classification and if it is 0, the email is valid and if is 1, it is the spam. Then a fuzzy tree is built from the training data and the results are reinvestigated.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

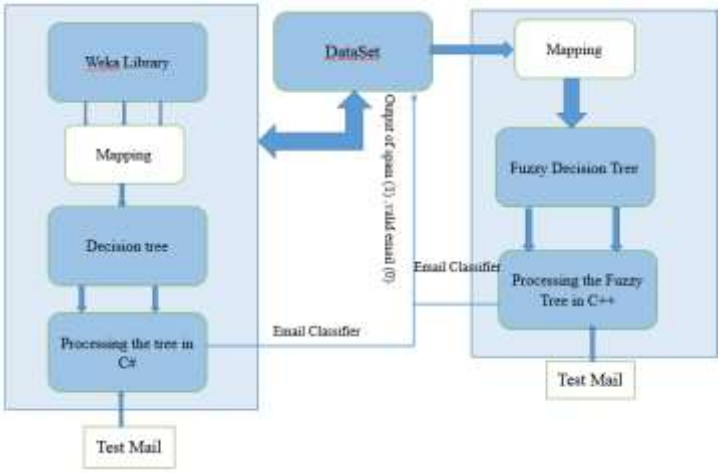

Figure.1. Architecture of spam filtering by the help of FSF.

C. Decision Tree Building

The decision tree is built by Weka Software and converted into XML format and processed in C# programming language and 500 test Emails are given to it and the precision of tree examined; the stages ate presented in details as follows:

Figure 2 shows a part of j48 tree by the output of Weka Software.

```
word_freq_hp <= 0.11
|   word_freq_george <= 0.2
|   |   word_freq_meeting <= 0.49
|   |   |   capital_run_length_longest <= 9
|   |   |   |   word_freq_remove <= 0.02
|   |   |   |   |   word_freq_free <= 0.14
|   |   |   |   |   |   word_freq_business <= 0.07
|   |   |   |   |   |   |   char_freq_! <= 0.809
|   |   |   |   |   |   |   |   word_freq_email <= 0.42
|   |   |   |   |   |   |   |   |   word_freq_project <= 0.28
|   |   |   |   |   |   |   |   |   |   word_freq_your <= 0.51
|   |   |   |   |   |   |   |   |   |   |   word_freq_mail <= 0.08: 0 (168.0/13.0)
|   |   |   |   |   |   |   |   |   |   |   word_freq_mail > 0.08
|   |   |   |   |   |   |   |   |   |   |   |   word_freq_you <= 1.68: 0 (6.0)
|   |   |   |   |   |   |   |   |   |   |   |   word_freq_you > 1.68: 1 (3.0)
|   |   |   |   |   |   |   |   |   |   word_freq_your > 0.51
|   |   |   |   |   |   |   |   |   |   |   word_freq_internet <= 0.22
|   |   |   |   |   |   |   |   |   |   |   |   capital_run_length_longest <= 7
|   |   |   |   |   |   |   |   |   |   |   |   |   word_freq_your <= 1.68
|   |   |   |   |   |   |   |   |   |   |   |   |   |   capital_run_length_average <= 1.382: 1 (11.0/4.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   capital_run_length_average > 1.382: 0 (8.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   word_freq_your > 1.68: 0 (19.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   capital_run_length_longest > 7: 1 (2.0)
|   |   |   |   |   |   |   |   |   |   |   word_freq_internet > 0.22: 1 (3.0)
|   |   |   |   |   |   |   |   |   word_freq_project > 0.28: 0 (13.0)
|   |   |   |   |   |   |   |   word_freq_email > 0.42
|   |   |   |   |   |   |   |   |   word_freq_you <= 3.15
|   |   |   |   |   |   |   |   |   |   capital_run_length_longest <= 5: 0 (5.0)
|   |   |   |   |   |   |   |   |   |   capital_run_length_longest > 5: 1 (3.0/1.0)
|   |   |   |   |   |   |   |   |   word_freq_you > 3.15: 1 (5.0)
```

Figure.2. A part ofj48 tree produced by Weka.

Figure 3 summarizes the results of j48 classifier in Weka including the precision and the number of data which is correctly or incorrectly classified.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

Correctly Classified Instances	2769	94.2157 %
Incorrectly Classified Instances	170	5.7843 %
Kappa statistic	0.8774	
Mean absolute error	0.0778	
Root mean squared error	0.2311	
Relative absolute error	16.4527 %	
Root relative squared error	47.5321 %	
Total Number of Instances	2939	

Number of Leaves : 59

Size of the tree : 117

Figure.3. Summary of j48 classifierinWeka.

Figure 4 shows a part of j48 tree by the output of Weka Software.

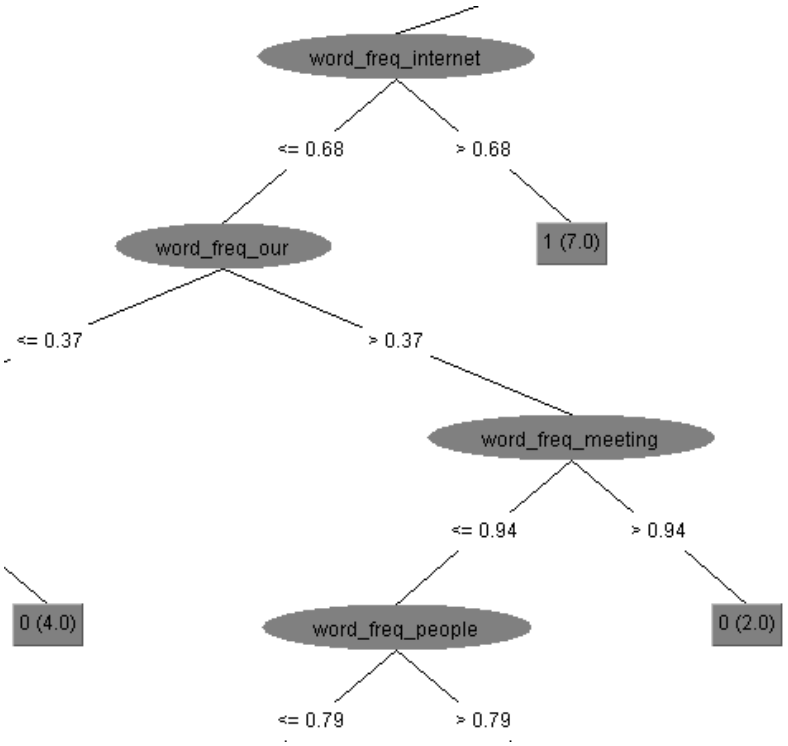

Figure.4.a part of j48 tree in Weka.

D. Building the fuzzy decision tree

A fuzzy decision tree combines fuzzy sets with symbolic decision trees. This fusion enhances the representative power of decision trees with the knowledge component inherent in fuzzy logic, leading to better robustness (noise immunity) and applicability in imprecise context [Janikow et al 2005][Mitra et al 2002]. Each fuzzy decision tree has three major components:

1. to partition each attribute into fuzzy sets and to assign membership degrees to its original values according to membership functions.

2. to use assigned fuzzy membership to induce an explicit fuzzy decision tree;

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

3. to infer from the tree in order to classify an instance.

In this section we use fuzzy ID3 [Maher et al 1993] [Umanol et al 1994], are preventative popular fuzzy decision tree for classifying static data, as a vehicle of illustration to explain these three components in detail.

D.1. Fuzzy partitioning of attributes

In the fuzzy ID3 system, fuzzy sets and membership degrees for every attribute isprovided by users.

D.2. Growing method

Regarding its growing method, FID3 is very similar to the crisp decision tree ID3[Quinlan 1993] except that FID3 uses membership degrees of instances rather than their crisp values to compute the information gain at each node. Assume that we have a fuzzy set S data with p attributes $x_1, \ldots, x_p$a class label in $\{c_1, \ldots, c_q\}$, andfuzzy sets $F_1, \ldots, F_m$ for each attribute x_j. Let S^{c_k} be the fuzzy subset in S whoseclass is c_k and let |S| be the sum of the membership degrees in the fuzzy setS. Then, the second component to build the FID3 is shown in Algorithm1.

The information gain $G(x_j, S)$ of the attribute $x_j$in the fuzzy set Sis defined by

$$G(x_j, S) = I(S) - E(x_j, S) \qquad (2)$$

Where

$$I(S) = -\sum_{k=1}^{q} (P(c_k, S).\, log_2^{P(c_k,S)}) \qquad (3)$$

$$E(x_j, S) = \sum_{v=1}^{m} P(S_{jv}, S)\ .\mathrm{I}(S_{jF_v})) \qquad (4)$$

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

Algorithm 1 FID3 learning [Umanol et al 1994]

1: Generate the root node with a fuzzy set comprising all instances with the membership degree 1;
2: **if** a node t with a fuzzy set of instances S satisfies one of the following conditions:
(**a**) the proportion of instances of a class c_k is greater than or equal to a threshold τ_1, that is, $\frac{S^{c_k}}{|S|} \geq \tau_1$;
(**b**) the number of instances is less than a threshold τ_2, that is, $|S| < \tau_2$;
(**c**) there are no attributes for further splitting;
then
3: The node t is a leaf and is assigned class labels with membership degrees;
4: **else**
5: **for** each $\mathbf{x}_j (j = 1, \ldots, p)$ **do**
6: Calculate the information gain $G(\mathbf{x}_j, S)$ following Eq.2;
7: Select the test attribute $\mathbf{x}_{max}$ that maximizes the information gain;
8: **end for**
9: Divide S into fuzzy subsets $S_1, \ldots, S_m$ according to $\mathbf{x}_{max}$, where the membership degree of every instance $\mathbf{x}^i$ in S_v is the product of $\mathbf{x}^i$'s membership degree in S and the value of $\mu_{max,F_v}(\mathbf{x}^i)$;
10: **for** each fuzzy subset $S_v (v = 1, \ldots, m)$ **do**
11: Generate a new node t_v corresponding to S_v;
12: Label the membership functions $\mu_{max,v}$ on the edge that connects t_v to t;
13: Replace S by S_v;
14: Repeat from Line 2 recursively;
15: **end for**
16: **end if**

Figure.5. Algorithm FID3 learning

$$P(c_k, S) = \frac{|S^{c_k}|}{|S|} \tag{5}$$

$$P(S_{jv}, S) = \frac{|S_{jF_v}|}{\sum_{v=1}^{m} |S_{jF_v}|} \tag{6}$$

If at least one of the conditions (a), (b)or(c) at Line 2 of Algorithm1is satisfied, the algorithm refrains from further splitting the node, and the node is assigned class labels together with membership degrees. This is in contrast to traditional decision trees that assign just one class label to each leaf. For assigning each membership degree$\beta_{kl} (0 \leq \beta_{kl} \leq 1)$ to the kth class at the lth leaf , FID3 takes into account all data:

$$\beta_{kl} = \frac{\sum_{i=1}^{n} \prod_{j \in path_l} \mu_{jl}(x^i) \mu_{kl}(y^i)}{\sum_{i=1}^{n} \prod_{j \in path_l} \mu_{jl}(x^i)} \tag{7}$$

Where $path_l$ is composed of the attributes that x^i has past when traversing from the rootnode to the lth leaf , and $\mu_{kl}(y^i)$ is the membership degree of x^iin the class k at the lth leaf.

D.3. Classifying new instances

When an unlabeled instance arrives to be classified, it traverses the tree from the root downward to the leaves of the tree. If the instance has a nonzero membership in at least one of the fuzzy sets associated with a node, the membership degree(s) is calculated and the instance is sent along the path (es) corresponding to the fuzzy set(s). The process is followed

(8)

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

recursively such that the instance meets all probable leaves. Thethird component, to calculate the probability of the instance x^i belonging to the classk, sums up the probability values across all leaves [Hashemi et al 2009]:

$$\widehat{y_k^i} = \sum_{\ell=1}^{l} \mu_{path_\ell}(x^i) \times \beta_{kl}$$

Where $0 \leq \widehat{y_k^i} \leq 1$ $(k = 1, \dots, q)$. Then, the final class label estimated by the learning model for the x^i) is $\hat{y}_i = arg\ max\ \ \{\widehat{y_k^i}\}_{k=1\dots q}$.

IV. Results And Analysis

From the decision trees, we selected j48 tree with the better result and compared the fuzzy tree and its result with the most commonly used techniques for spam filtering, SVM and Naïve Bayesian. As mentioned, we applied the spam base including 4601 including 1813(39.4%) spam emails and 2788 (60.6%) valid emails. Furthermore, we utilized 4101 emails for training and 500 ones for test.

The target tree is built from the training data by the help of Weka Software, and then the tree is processed and the test emails are given to it; this procedure is repeated for fuzzy tree and they are compared with the test results of SVM and Naïve Bayesian:

The definitions of mentioned items in the following tables and diagrams:

TP: The spam which is correctly detected as the spam; in other words, a record of test data with the original class 1 is put in the Class 1 by the classifier.

FP: The valid email which is predicted as the spam by mistake; in other words, a record of test data with the original class 0 is put in the Class 1 by the classifier.

TN: The valid email which is predicted as the valid email; in other words, a record of test data with the original class 0 is put in the Class 0 by the classifier.

FN: The spam email which is predicted as the valid email by mistake; in other words, a record of test data with the original class 1 is put in the Class 0 by the classifier.

F-Measure :weighted average of the precision and recall.

$$Precision - Class\ 1 = \frac{TP}{(TP+FP)} \quad (9)$$

$$Recall - Class\ 1 = \frac{TP}{(TP+FN)} \quad (10)$$

$$F - Measure = \frac{Precision*Recall}{Precision+Recall} \quad (11)$$

$$Precision - Class\ 0 = \frac{TN}{(TN+FN)} \quad (12)$$

$$Recall - Class\ 0 = \frac{TN}{(TN+FP)} \quad (13)$$

A. Evaluation of results

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

The target tree is built from the training data by the help of Weka Software, and then the tree is processed and the test emails are given to it and the result shown in table 1 and 2.

A.1. Evaluating the results of spam class:

TABLE1

RESULTS OBTAINED FROM THE PROCESSING OF SPAM CLASS

	TP	FP	Precision	Recall	F-Measure
SVM	0.757	0.024	0.959	0.757	0.846
Naive Bayesian	0.963	0.231	0.757	0.963	0.848
J48	0.888	0.049	0.931	0.888	0.909
FID3	0.897	0.076	0.897	0.941	0.918

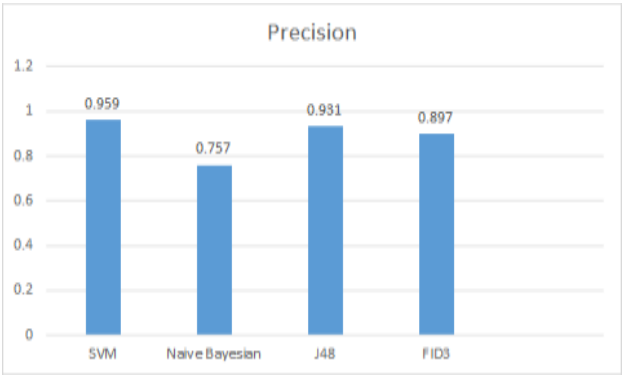

Figure.6. Comparison of precision, The results obtained from processing the spam class with involving all features.

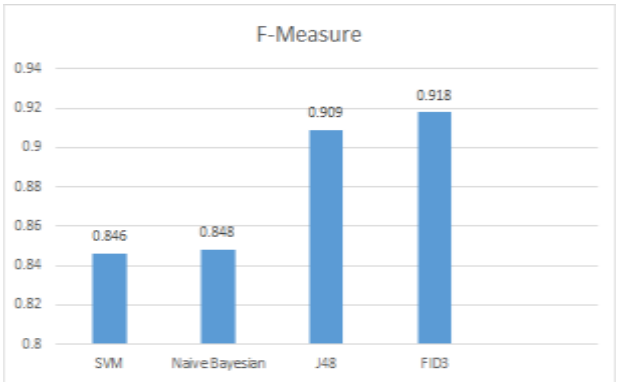

Figure.7. Comparison of F-Measure, The results obtained from processing the spam class with involving all features

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

As shown in Figure 7, the F-Measure of FID3 Fuzzy Tree is higher than others indicating the balance between both classes unlike SVM and Naïve Byesian.

A.2. Evaluating the results in the valid email class:

TABLE1

RESULTS OBTAINED FROM THE PROCESSING THE CLASS OF VALID EMAIL

	TN	FN	Precision	Recall	F-Measure
SVM	0.976	0.243	0.843	0.976	0.904
Naive Bayesian	0.769	0.037	0.965	0.769	0.856
J48	0.951	0.112	0.919	0.951	0.935
FID3	0.958	0.056	0.944	0.925	0.941

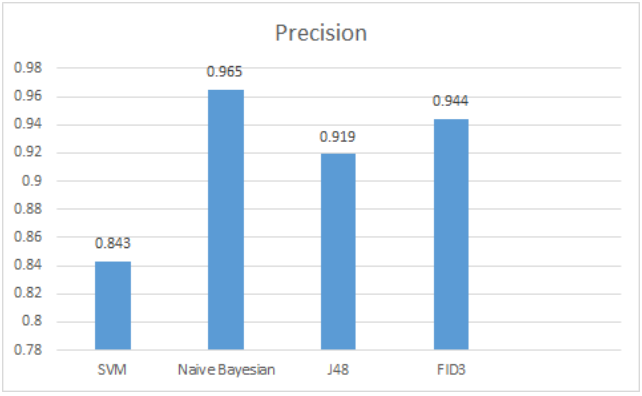

Figure.8. Comparison of precision, The results obtained from processing the valid email class.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

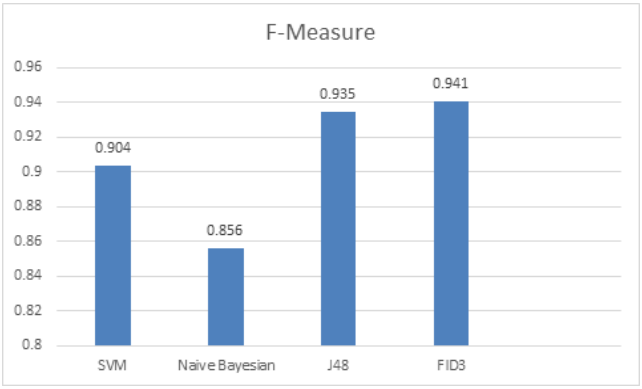

Figure.9. Comparison of F-measure, The results obtained from processing the valid email class.

As observed, the best precision in the class of valid email is related to Naïve Bayesian classifier, but if the results of both classes are investigated for different classifiers, a more comprehensive conclusion is obtained. Figure 10 clearly shows this issue.

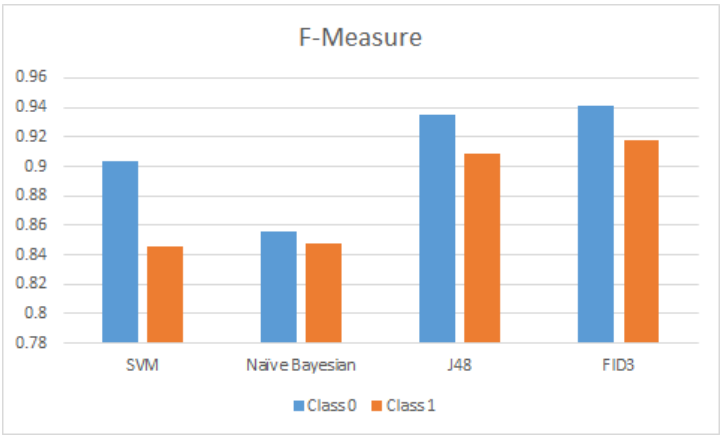

Figure.10. Results of both classes

As shown in Figure 10, our method is able to maintain an acceptable balance between errors of both classes.

V. Conclusion and future works

The important objective of this research is to utilize FID3 fuzzy decision tree for classification of emails, thus the fuzzy decision trees are built from the training data and applied for classification of email; finally, we were able to make balance between two errors as the result of detecting the valid email instead of spam and vise versa.

Pruning and slimming the generated tree can be considered as the future works. Furthermore, other non-tree classifiers and decision trees such as AD, LAD can be appliedin addition toFID3 fuzzy decision tree in order to achieve higher precision; moreover, feature selection

ISSN 1925-931X

Available online @www.academians.org

and dimension reduction methods such as tf-idf and Binomial hypothesis testing can be utilized in this regard.

VI. References

Blachnik, M., Duch,W., Kachel, A. and Biesiada, J (2009). "Feature Selection for Supervised Classification: A Kolmogorov-Smirnov Class Correlation-Based Filter," in *AIMeth, Symposium on Methos of Artificial Intelligence, Gliwice, Poland*.

Chang, M. and C. K. Poon (2009). "Using phrases as features in email classification," *Journal of Systems and Software*, vol. 82, pp. 1036-1045.

Çıltık, A. and Güngör, T (2008). "Time-efficient spam e-mail filtering using< i> n</i>-gram models," *Pattern Recognition Letters*, vol. 29, pp. 19-33.

Goodman, J (2004). IP Address in Email Clients, Conf on Email and Anti-pam (CEAS), California, USA, July 30-31.

Hashemi, S. & Yang, Y. (2009). Flexible decision tree for data stream classification in the presence of concept change, noise and missing values. *Data Mining and Knowledge Discovery*, *19*(1), 95-131.

Hu, H., Yu, B (2010). Automatic Thesaurus Construction for Spam Filtering using Revised Back Propagation Neural Network, Expert system with application, vol. 37, no. 1, pp. 18-23.

Janikow, C.Z. and Kawa, K (2005). "Fuzzy decision tree FID," in *Fuzzy Information Processing Society, Annual Meeting of the North American*, pp. 379-384.

Lai, C.C(2007). An Empirical Study of Three Machine Learning Methods for Spam Filtering, Knowledge-Based systems, vol. 20, no. 3, pp. 249-254.

MAAWG.(2006).Messaging anti-abuse working group.Email metrics report.Third& fourth guarter.Available at http://www.maawg.org/about/MAAWGMetric2006 3 4 report.pdf Accessed: 04.06.07.

Maher, P.E and St Clair, D (1993). "Uncertain reasoning in an ID3 machine learning framework," in *Fuzzy Systems, Second IEEE International Conference on*, pp. 7-12.

Mitra, S., Konwar, K.M. and Pal, S.K. (2002). "Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 32, pp. 328-339.

Quinlan, JR (1993). C4.5: Programs for machine learning. Morgan Kaufmann Publishers Induction of decision trees, pp 349–361

spambasedadaset available in: http://archive.ics.uci.edu/ml/datasets/Spambase.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 4(1) January 2014, pp. 38-50

Int. Association Academians

ISSN 1925-931X

Available online @www.academians.org

Umanol, M., Okamoto, H., Hatono, I., Tamura, H., Kawachi, F., Umedzu, S., and Kinoshita, J. (1994). "Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems," in *Fuzzy Systems, IEEE World Congress on Computational Intelligence., Proceedings of the Third IEEE Conference on*, pp. 2113-2118.

Wang, B. et al, (2006).Using Online Linear Classifiers to Filter Spam Emails, Pattern Analysis & Application, vol. 9, no. 4, pp. 339-351.

Watson, B, (2004). "Beyond Identity: Addressing Problems that Persist in an Electronic Mail System with Reliable Sender Identification," in *CEAS*.