Journal of Information Hiding and Multimedia Signal Processing ©2022 ISSN 2073-4212
Ubiquitous International Volume **13**, Number **4**, December **2022**

Assessing some formula of spam probability and applying in spam classification

[1]Trung Nguyen Tu*, [2]Nguyen Thi Loan, [3]Vu Thi Khanh Toan

[1]Faculty of Information Technology, Thuyloi University, 175 Tay Son, Hanoi, Vietnam
[2]Hanoi Pedagogical University 2, 32 Nguyen Van Linh, Xuan Hoa, Phuc Yen, Vinh Phuc, Vietnam
[3]Vietnam National University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam
Email: trungnt@tlu.edu.vn, nguyenthiloan@hpu2.edu.vn, vtktoan@vnua.edu.vn

Received April 2022; revised October 2022

Abstract. *Spam classification is a problem that has been studied for a long time in the world. The spam classification feature is integrated into the Mail Server or Mail Client. Traditional methods still have certain weaknesses, the content-based classification method proved effective with the use of techniques in statistical machine learning. In particular, spam classification based on Bayes with the advantages of simplicity, ease of use and fast speed. This article presents an assessment of some methods to calculate the probability of being Spam of Tokens through spam classification application.*

Keywords: Spam, Ham, Spam classification, Spam probability, Tokens.

1. **Introduction.** One of the services that the Internet provides is email service. It is a very simple, convenient, cheap and effective means of communication between people in the community using Internet services. However, Due to the benefits of email services that the number of messages exchanged on the Internet is increasing, and most of them are spam. Spam mail is unsolicited, unsolicited and mass email messages sent to recipients. Spam is often sent in very large numbers, not expected by users, often with the purpose of advertising, attaching viruses, causing discomfort to users, reducing internet transmission speed and processing speed of users. email server, causing huge economic losses. According to the statistics of kaspersky in 2014 [12], the percentage of spam in email traffic in February increased by 4.2There are different methods of spam filtering. Each method has its own advantages and disadvantages. In particular, the content filtering method to classify spam has been the most interested, researched and applied method. This method relies on the body and subject of the message to distinguish spam from legitimate messages. This method has the advantage that we can easily change the filter so that it can filter spam types accordingly. In content-based learning, spam filtering using statistical machine learning techniques is a promising method for many commercial applications such as Hotmail, Google, Yahoo. Machine learning methods and statistical probabilities allow the classification of spam that has never appeared before. In [1], Awad presented an evaluation, comparing several machine learning methods (Bayesian classification, k-NN, ANNs, SVMs...) for spam filtering problem. In [6], Shahar Yifrah and Guy Lev present a project to build a spam filter using machine learning techniques. In [10], the authors compared the effectiveness of different spam filters using Naïve Bayes, SVM, and KNN. The test results show that filters using these techniques give very high accuracy. The peculiarity of content-based techniques is to analyze the word in the content and calculate the token or feature value. Once the number of tokens and features is large, methods

like SVMs, ANNs have a very slow training speed. Among the spam filtering techniques based on statistical machine learning, Bayesian technique proves to be simple, effective, and has a very fast execution speed, not only in the classification stage but also during training. Bayesian algorithm has been applied to spambayes spam filtering program, and the filtering results are quite effective. Perhaps, this is the reason that filters using this technique are commonly installed in Mail Server (Zimbra) and Mail Client systems. Mail Client software such as Outlook, Outlook Express, Thunderbird/Mozilla Mail & Newsgroups, Eudora, or Opera Mail. Naïve Bayes algorithms are the classic algorithms in Bayesian engineering. Naïve Bayes is very popular among open source anti-Spam email filters [9]. There are many versions of Naïve Bayes. In [9], the authors discussed, tested and evaluated the spam filtering efficiency of these versions. In [5], Phan Huu Tiep and his colleagues present the Vietnamese spam filtering process based on the Naïve Bayes algorithm and the processing of Vietnamese sentence separation. In [7], Tianda et al. presented a comparison between a spam classifier using only Naïve Bayes techniques and a spam classifier using a technical spam classifier and association rules. In [4], the authors discuss a statistical spam filtering process using the Naïve Bayes classification technique. A convenient, simple way to implement the Bayesian algorithm in spam filtering is the algorithm by Paul Graham [8][4] and another variation by Tim Peter. These algorithms all analyze, evaluate and make suggestions on ways to calculate the spam probability of tokens. In it, Paul Graham's improvement gives very high accuracy. In [2], Jialin et al discussed and evaluated the spam SMS filtering method using SVM and MTM (message topic model). Bayesian network is also widely used in recent years [14-17]. The Bayesian spam classification method (especially by Paul Graham) has a number of limitations including: (1) not fully considering the factors affecting each token and (2) some unresolved cases good. This article evaluates some ways to calculate the Spam probability of tokens from analyzing Paul Graham's Spam probability formula and improving the token's Spam probability formula. The next sections are presented as follows. Part 2 presents the problem of Bayes-based spam filtering. Part 3 presents several different ways of calculating the Spam probability of tokens. The trials are presented in section 4. Conclusions are presented in section 5.

2. **Related Work.**

2.1. **Classifying Spam using Bayes.** Bayesian spam classification technique is presented in [3][5]. Consider each email represented by a feature vector $\vec{x} = (x_1, x_2, ..., x_n)$ with $x_1, x_2, ..., x_n$ they are the values of the attributes $X_1, X_2, ..., X_n$ corresponds in the feature space. Using binary values 0 and 1 to describe that email has the feature X_i or not. Assume that the email has the feature X_i, setting the value of $X_i = 1$. Otherwise, setting the value of $X_i = 0$.

From Bayesian probability theory we have the formula for the probability of mail with vector $\vec{x} = (x_1, x_2, ..., x_n)$ belongs to class c as follows:

$$P\left(C = c|\vec{X} = \vec{x}\right) = \frac{P\left(C = c\right) P\left(\vec{X} = \vec{x}|C = c\right)}{\sum_{k\in\{Spam,Ham\}} P\left(C = k\right) P\left(\vec{X} = \vec{x}|C = c\right)} \tag{1}$$

For simplicity when calculating $P\left(\vec{X}|C\right)$, we have to assume $X_1, X_2, \ldots, X_n$ independent. Therefore, the expression (1) is equivalent to the following expression:

$$P\left(C=c|\vec{X}=\vec{x}\right)=\frac{P\left(C=c\right)\prod_{i=1}^{n}P\left(X_i=x_i|C=c\right)}{\sum_{k\in\{Spam,Ham\}}P\left(C=k\right)\prod_{i=1}^{n}P\left(X_i=x_i|C=c\right)} \tag{2}$$

The most widely used value to rank an attribute is a mutual value MI (mutual information). The mutual value MI for which each representation of X belongs to type C is calculated as follows:

$$MI=\sum_{x\in 0,1,c\in Spam,Ham}P\left(X=x|C=c\right)log\frac{P\left(X=x|C=c\right)}{P\left(X=x\right)P\left(C=c\right)} \tag{3}$$

An email is considered spam if:

$$\frac{P\left(C=Spam|\vec{X}=\vec{x}\right)}{P\left(C=Ham|\vec{X}=\vec{x}\right)}>\lambda \tag{4}$$

Where λ is a given threshold to consider comparing with the ratio between the probability of being Spam or Ham of a message. In which, Spam: spam, Ham: valid mail.

Assume the attributes X_i is independent. Then, we have:

$$P\left(C=Spam|\vec{X}=\vec{x}\right)=1-P\left(C=Ham|\vec{X}=\vec{x}\right) \tag{5}$$

Therefore, (4) is equivalent to:

$$P\left(C=Spam|\vec{X}=\vec{x}\right)>t \tag{6}$$

with $t=\frac{\lambda}{1+\lambda}$

2.2. **Formula of Paul Graham.** According to [8][4], Paul Graham proposed a way to calculate the spam probability of tokens. Paul Graham's formula is very simple, convenient for installation, but also for high spam classification accuracy. The formula for calculating Spam probability of token w as follows:

$$P\left(S|w\right)=\frac{\frac{SA(w)}{STM}}{\frac{SA(w)}{STM}+2\frac{HA(w)}{HTM}} \tag{7}$$

Where:

- SA(w): number of occurrences of token w in spam store.
- HA(w): number of occurrences of token w in valid message store.
- STM: total number of messages in spam store.
- HTM: total number of messages in valid mail store.

Factor "2" to increase the likelihood of receiving a valid message.

Training dataset in [4] includes 432 spam and 2170 ham [4].

At this time, the probability of being Spam of a message E is calculated by the formula:

$$P\left(S|E\right)=\frac{\prod_{i=1}^{n}P\left(S|w_i\right)}{\prod_{i=1}^{n}P\left(S|w_i\right)+\prod_{i=1}^{n}P\left(H|w_i\right)} \tag{8}$$

Where:

$$P\left(H|w_i\right)=1-P\left(S|w_i\right) \tag{9}$$

Table 1. Training data table in [4].

Token	Number in Spam	Number in Ham	P $(S\|w)$
A	165	1235	0.2512473
Advised	12	42	0.4177898
As	2	579	0.0086009
Chance	45	35	0.7635468
Clarins	1	6	0.2950775
Exercise	6	39	0.2787054
For	378	1829	0.3417015
Free	253	137	0.8226372
Fun	59	9	0.9427419
Girlfriend	26	8	0.8908609
Have	291	2008	0.2668504
Her	38	118	0.4471509
I	9	1435	0.0155078
Just	207	253	0.6726596
Much	126	270	0.5396092
Now	221	337	0.6222218
Paying	26	10	0.8671995
Receive	171	98	0.8142107
Regularly	9	87	0.2062346
Take	142	287	0.5541010
Tell	76	89	0.6820062
The	185	930	0.3331618
Time	212	446	0.5441787
To	389	1948	0.3340176
Too	56	141	0.4993754
Trial	26	13	0.8339739
Vehicle	21	58	0.4762651
Viagra	39	19	0.8375393
You	391	786	0.5554363
Your	332	450	0.6494897

3. **Some improvements in token's spam probability calculation.** From formula (7), we have the following observations:

- Calculating the probability of being Spam of each token
  - Depends only on the number of occurrences of token w and the total number of messages in each spam and valid dataset.
  - Doesn't consider the total frequency of all tokens.
  - Do not consider the number of messages containing tokens in each spam and valid mail vault. At this time, it is not known whether the token appears in only one message or many messages.
  - A factor of "2" increases the chances of misrepresenting spam as valid mail. This is very dangerous if the message contains a virus because if it is a valid message, the user will be "safer" than clicking on the message.
- In case the number of occurrences of a certain token is approximately or equal to the total number of messages in the spam vault and appears very little in the valid archive. At this time, the "SA(w)/STM" ratio will be close to or equal to 1 while

the "HA(w)/HTM" ratio will gradually approach 0. There is a probability that the Spam of token w will accordingly approach or equal to 1 (according to formula 7). From here, according to formula (8), the probability that the Spam of a message containing this token will be very high or equal to 1. In other words, the probability that the Spam of a message containing this token will be affected almost exclusively by this token. For example, if a message appears with this token only once, the other tokens in this message have a low probability of spam but this message is considered very Spam. This is unreasonable.

From the above analysis, we find the following: The probability of being Spam of each token can depend on the following factors:

a) umber of occurrences of token w in each store: spam and valid mail.

b) Total number of messages in each store: spam and valid mail.

c) Total frequency of all tokens.

d) Number of messages containing tokens in each store: spam and valid mail.

In addition, changing the factor "2" in different cases to enhance the ability to recognize spam or legitimate mail.

From here, we give some formulas to calculate the probability of being Spam of each token as follows:

- Depending on factors a-c, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}}{\frac{SA(w))}{STA}+\frac{HA(w))}{HTA}} \tag{10}$$

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}}{\frac{SA(w))}{STA}+2\frac{HA(w))}{HTA}} \tag{11}$$

$$P\left(S|w\right)=\frac{2\frac{SA(w))}{STA}}{2\frac{SA(w))}{STA}+\frac{HA(w))}{HTA}} \tag{12}$$

- Depending on factors a-b, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}}{\frac{SA(w))}{STM}+\frac{HA(w))}{HTM}} \tag{13}$$

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}}{\frac{SA(w))}{STM}+2\frac{HA(w))}{HTM}}(PaulGraham) \tag{14}$$

$$P\left(S|w\right)=\frac{2\frac{SA(w))}{STA}}{2\frac{SA(w))}{STM}+\frac{HA(w))}{HTM}} \tag{15}$$

- Depending on factors b-d, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{STM(w))}{STM}}{\frac{STM(w))}{STM}+\frac{HTM(w))}{HTM}} \tag{16}$$

$$P\left(S|w\right)=\frac{\frac{STM(w))}{STM}}{\frac{STM(w))}{STM}+2\frac{HTM(w))}{HTM}} \tag{17}$$

$$P\left(S|w\right)=\frac{2\frac{STM(w))}{STM}}{2\frac{STM(w))}{STM}+\frac{HTM(w))}{HTM}} \tag{18}$$

- Depending on factors c-d, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STM(w)}}{\frac{SA(w))}{STM(w)}+\frac{HA(w))}{HTM(w)}} \tag{19}$$

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STM(w)}}{\frac{SA(w))}{STM(w)}+2\frac{HA(w))}{HTM(w)}} \tag{20}$$

$$P\left(S|w\right)=\frac{2\frac{SA(w))}{STM(w)}}{2\frac{SA(w))}{STM(w)}+\frac{HA(w))}{HTM(w)}} \tag{21}$$

- Depending on factors a-b-d, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STM}*\frac{STM(w))}{STM}}{\frac{SA(w))}{STM}*\frac{STM(w))}{STM}+\frac{HA(w))}{HTM}*\frac{HTM(w))}{HTM}} \tag{22}$$

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STM}*\frac{STM(w))}{STM}}{\frac{SA(w))}{STM}*\frac{STM(w))}{STM}+2\frac{HA(w))}{HTM}*\frac{HTM(w))}{HTM}} \tag{23}$$

$$P\left(S|w\right)=\frac{2\frac{SA(w))}{STM}*\frac{STM(w))}{STM}}{2\frac{SA(w))}{STM}*\frac{STM(w))}{STM}+\frac{HA(w))}{HTM}*\frac{HTM(w))}{HTM}} \tag{24}$$

- Depending on factors a-b-c-d, we get the formulas:

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}*\frac{STM(w))}{STM}}{\frac{SA(w))}{STA}*\frac{STM(w))}{STM}+\frac{HA(w))}{HTA}*\frac{HTM(w))}{HTM}} \tag{25}$$

$$P\left(S|w\right)=\frac{\frac{SA(w))}{STA}*\frac{STM(w))}{STM}}{\frac{SA(w))}{STA}*\frac{STM(w))}{STM}+2\frac{HA(w))}{HTA}*\frac{HTM(w))}{HTM}} \tag{26}$$

$$P\left(S|w\right)=\frac{2\frac{SA(w))}{STA}*\frac{STM(w))}{STM}}{2\frac{SA(w))}{STA}*\frac{STM(w))}{STM}+\frac{HA(w))}{HTA}*\frac{HTM(w))}{HTM}} \tag{27}$$

If the formulas 10-27 is used, the problem in comment (2) can be overcome.

4. **Experiments.** The sample data set is CSDMC2010_SPAM [11]. The training data set includes SpamTrain and HamTrain.

4.1. **Expriment 1.** HamTrain has 2808 valid mails, SpamTrain has 1238 spam. The test data set includes HamTest (141 valid mails) and SpamTest (140 spam). Tables 2, 3 and 4 statistics the accuracy of Spam classification through Precision index statistics in cases: no factor "2", factor "2" to increase classification into valid mail, factor "2" to enhance classification into spam.

From Table 2, we see that the SPAM classification accuracy of the formulas 12, 16 and 22 is the highest. Meanwhile, the HAM classification accuracy of the 10 formulas is the highest.

From Table 3, we see that the SPAM classification accuracy of the formulas 23 is the highest. Meanwhile, the HAM classification accuracy of the formulas 11 and 14 is the highest.

From Table 4, we see that the SPAM classification accuracy of the formulas 15, 18 and 24 is the highest. Meanwhile, the HAM classification accuracy of the 12 formulas is the highest.

Table 2. Statistics on classification accuracy of spam and valid mail in the absence of a factor of 2.

Formula	SPAM	HAM
10	62.857	96.454
13	98.571	92.908
16	98.571	90.780
19	90.714	94.326
22	98.571	85.816
25	94.286	92.199

Table 3. Spam and valid mail collection classification accuracy statistics in the case of a factor of 2 to increase classification to valid mail.

Formula	SPAM	HAM
11	83.571	96.454
14	89.286	96.454
17	87.143	95.035
20	82.143	95.745
23	93.571	92.908
26	80.714	93.617

Table 4. Statistics on classification accuracy of spam and valid mail in the case of a factor of 2 to increase classification as spam.

Formula	SPAM	HAM
12	97.857	92.908
15	99.286	82.270
18	99.286	80.142
21	98.571	85.816
24	99.286	79.433
27	98.571	86.525

4.2. **Experment 2.** HamTrain has 2535 valid mails, SpamTrain has 1014 spam. The test data set includes HamTest (414 valid mails) and SpamTest (364 spam). Tables 5, 6 and 7 statistics the accuracy of Spam classification through Precision index statistics in cases: no factor "2", factor "2" to increase classification into valid mail, factor "2" to enhance classification into spam.

Table 5. Statistics on classification accuracy of spam and valid mail in the absence of a factor of 2.

Formula	SPAM	HAM
10	59.066	98.068
13	98.077	95.652
16	98.626	93.720
19	89.835	96.135
22	98.901	87.923
25	93.132	93.237

From Table 5, we see that the SPAM classification accuracy of the formulas 22 is the highest. Meanwhile, the HAM classification accuracy of the 10 formulas is the highest.

Table 6. Spam and valid mail collection classification accuracy statistics in the case of a factor of 2 to increase classification.

Formula	SPAM	HAM
11	78.571	97.826
14	86.813	98.068
17	88.736	96.618
20	77.747	97.826
23	90.659	93.720
26	77.473	94.686

From Table 6, we see that the SPAM classification accuracy of the formulas 23 is the highest. Meanwhile, the HAM classification accuracy of the formulas 14 is the highest.

Table 7. Statistics on classification accuracy of spam and valid mail in the case of a factor of 2 to increase classification as spam.

Formula	SPAM	HAM
12	95.879	94.686
15	99.725	84.541
18	99.725	82.126
21	98.626	87.923
24	99.725	81.159
27	98.077	89.855

From Table 7, we see that the SPAM classification accuracy of the formulas 15, 18 and 24 is the highest. Meanwhile, the HAM classification accuracy of the 12 formulas is the highest.

4.3. **Experment 3.** HamTrain has 2448 valid mails, SpamTrain has 986 spam. The test data set includes HamTest (501 valid mails) and SpamTest (392 spam). Tables 8, 9 and 10 statistics the accuracy of Spam classification through Precision index statistics in cases: no factor "2", factor "2" to increase classification into valid mail, factor "2" to enhance classification into spam.

Table 8. Statistics on classification accuracy of spam and valid mail in the absence of a factor of 2.

Formula	SPAM	HAM
10	58.929	98.204
13	98.469	95.808
16	98.469	93.613
19	90.051	96.407
22	98.980	88.224
25	91.837	92.814

From Table 8, we see that the SPAM classification accuracy of the formulas 22 is the highest. Meanwhile, the HAM classification accuracy of the 10 formulas is the highest.

Table 9. Spam and valid mail collection classification accuracy statistics in the case of a factor of 2 to increase classification to valid mail.

Formula	SPAM	HAM
11	78.571	98.004
14	85.459	98.204
17	87.500	96.607
20	76.786	98.004
23	90.051	93.413
26	75.765	94.810

Table 10. Statistics on classification accuracy of spam and valid mail in the case of a factor of 2 to increase classification as spam.

Formula	SPAM	HAM
12	95.918	94.611
15	99.745	85.030
18	99.745	82.236
21	98.724	87.625
24	99.745	82.036
27	97.959	89.820

From Table 9, we see that the SPAM classification accuracy of the formulas 23 is the highest. Meanwhile, the HAM classification accuracy of the formulas 14 is the highest.

From Table 10, we see that the SPAM classification accuracy of the formulas 15, 18 and 24 is the highest. Meanwhile, the HAM classification accuracy of the 12 formulas is the highest.

Through the tests, we can make the following observations:

- In the absence of factor “2”, formulas 13, 16 and 22 give the highest SPAM classification accuracy; Formula 10 gives the highest HAM classification accuracy.
- In the case of a factor of “2” to enhance validation, Equations 23 give the highest SPAM classification accuracy; formula 14 gives the highest HAM classification accuracy.
- In the case of coefficient “2” to enhance garbage collection, formulas 15, 18 and 24 give the highest SPAM classification accuracy; formula 12 gives the highest HAM classification accuracy.

5. **Conclusions.** In this paper, we discussed and analyzed Spam filtering techniques using Bayes. From there, give some ways to calculate the probability of being Spam of the token. Testing has shown them to be good alternatives to Bayesian-based Spam filters in different situations. Depending on the specific purpose of the application: keep the important type of HAM or eliminate the dangerous SPAM, choose the corresponding formula. In the next study, we plan to output the new Spam probability formula for each token using fuzzy logic.

Acknowledgment. This work is partially supported by Department of Information Technology, Thuyloi university and Institute of Information Technology, Vietnamese Academy of Science and Technology. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Awad W.A. and ELseuofi S.M., Machine learning methods for spam e-mail classification, International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011, pp.173-184.

[2] Jialin ma, Yongjun zhang, Jinling liu, Intelligent SMS spam filtering using topic model, ieee international conference on intelligent networking and collaborative systems (incos), 2016.

[3] Johan Hovol, Naïve Bayes Spam filtering using Word-Position-Based attributes, Proceedings of the 15th NODALIDA conference, 2006, pp. 78–87.

[4] Paul Graham, Better Bayesian filtering. In Proceedings of the 2003 Spam Conference (http://spamconference.org/ proceedings2003.html), Cambridge, MA, 2003.

[5] Phan Huu Tiep, Vu Duc Lung, Cao Nguyen Thuy Tien, Lam Thanh Hien, Phuong phap loc thu rac ting Viet dua tren tu ghep va theo vet nguoi su dung, Hoi thao "Mot so van de chon loc cua Cong nghe thong tin va truyen thong", Can Tho, 2011.

[6] Shahar Yifrah và Guy Lev, Machine Learning Final Project Spam Email Filtering, ML Project, 2013.

[7] Tianda Yang, Kai Qian, Dan Chia-Tien Lo, Spam filtering using Association Rules and Naïve Bayes Classifier, IEEE International Conference on Progress in Informatics and Computing (PIC), 2015.

[8] Tianhao Sun, Spam Filtering based on Naïve Bayes Classication, May 2009.

[9] Vangelis Metsis, Ion And rout sopoulos and Georgios Paliouras, Spam Filtering with Naïve Bayes–Which Naïve Bayes?, CEAS2006-Third Conference on Email and Anti-Spam, Mountain View, California USA, July 27-28, 2006.

[10] Yun-Nung Chen, Che-An Lu, Chao-Yu Huang, Anti-Spam Filter Based on Naïve Bayes, SVM, and KNN model, AI term project group 14, 2009.

[11] http://csmining.org/index.php/spam-email-datasets-.html

[12] http://kaspersky.nts.com.vn/

[13] http://antoanthongtin.vn/

[14] Hui Wang, TianWang Dai, Kun Liu, XinXin Ru and YaLong Lou, Node Confidence Calculation Method Based on D-separation of Bayesian Network, Journal of Network Intelligence, Vol. 5, No. 4, pp. 166-178, November 2020.

[15] Lili Meng, Jingxiu Zong, Guina Sun, Jie Cheng, Jia Zhang and Mengchen Zhao, Bayesian Multi-Hypothesis Wyner-Ziv Video Coding, Journal of Information Hiding and Multimedia Signal Processing, Vol. 8, No. 2, pp. 478-485, March 2017.

[16] Fuquan Zhang, Gangyi Ding, Zijing Mao, Lin Xu and Xiaoyan Zheng, Bayesian Network for Motivation Classification in Creative Computation, Journal of Information Hiding and Multimedia Signal Processing, Vol. 8, No. 4, pp. 888-902, July 2017.

[17] Chaur-Heh Hsieh, Chung-Ming Kuo and Yao-Sheng Hsieh, Bayesian-Based Probabilistic Architecture for Image Categorization Using Macro- and Micro-Sense Visual Vocabulary, Journal of Information Hiding and Multimedia Signal Processing, Vol. 9, No. 6, pp. 1628-1638, November 2018.