西南交通大学学报
第 54 卷 第 3 期
2019 六 3 月
JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY
Vol.54 No.3
June 2019

ISSN - 0258-2724
DOI：10.35741/issn.0258-2724.54.3.6

Research article

# AN ANTI-SPAM DETECTION MODEL FOR EMAILS OF MULTI-NATURAL LANGUAGE

**Mazin Abed Mohammed [a,*], Salama A. Mostafa [b,*], Omar Ibrahim Obaid [c], Subhi R. M. Zeebaree [d], Mohd Khanapi Abd Ghani [e], Aida Mustapha [b], Mohd Farhan Md Fudzee [b], Mohammed Ahmed Jubair [b], Mustafa Hamid Hassan [b], Azizan Ismail [b], Dheyaa Ahmed Ibrahim [f], Fahad Taha AL-Dhief [g]**

[a] College of Computer Science and Information Technology, University of Anbar
Anbar, mazinalshujeary@uoanbar.edu.iq
[b] Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia
Johor, Malaysia, salama@uthm.edu.my, aidam@uthm.edu.my, farhan@uthm.edu.my, azizanis@uthm.edu.my, mohamed.a.jubair@gmail.com, mustafa.hamid.alani@gmail.com
[c] Department of Computer, College of Education, AL-Iraqia University
Baghdad, Iraq, alhamdanyomar23@gmail.com
[d] Duhok Polytechnic University and Tishk International University
Duhok, Iraq and Erbil, Iraq, subhi.rafeeq@dpu.edu.krd
[e] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia, khanapi@utem.edu.my
[f] Computer Engineering Techniques Department, Imam Ja'afar Al-Sadiq University
Baghdad, Iraq, dheyaa.ibrahim88@gmail.com
[g] Faculty of Electrical Engineering, Department of Communication Engineering, Universiti Teknologi Malaysia, Johor, Malaysia, fahadtaha37@yahoo.com

**Abstract**

The spam is one of the illegal and negative practices that involves the use of email services to send unsolicited emails such as phishing for the purpose of scamming which influences the reliability of email. Investigations have been conducted from various perspectives in order to examine this spam problem and how it affects society. In this regard, many studies have been carried out with the aim of studying the effect of spam activity on finance, economy, marketing, business and management, while other studies have focused on studying the influence of spam on security and privacy. Consequently, the literature affords various anti-spam methods that blocks or filters spam emails. This paper investigates the existing anti-spam methods, highlights some current problems and carries out an improved anti-spam model. In this regard, a new agent-based of Multi-Natural Language Anti-Spam (MNLAS) model is proposed. The MNLAS model process in the spam filtering process of an email both visual information such as images and texts in English and Arabic languages. The Jade agent platform and Java environments are employed in the implementation of MNLAS model. The MNLAS model is tested on a 200 emails' dataset and the results show that it is able to detect and filter various kinds of spam emails with high accuracy.

**Keywords:** Anti-spam classification, machine learning, software agent, multi-agent system.

**摘要：** 垃圾邮件是非法和消极行为之一，涉及使用电子邮件服务发送未经请求的电子邮件（如网络钓鱼），以防止影响电子邮件可靠性的欺诈行为。从各个角度进行调查，以检查这一垃圾邮件问题及其如何影响社会。在这方面，已经进行了许多研究，目的是研究垃圾活动对金融，经济，营销，商业和管理的影响，

而其他研究则侧重于研究垃圾邮件对安全和隐私的影响。因此，文献提供了阻止或过滤垃圾邮件的各种反垃圾邮件方法。本文研究了现有的反垃圾邮件方法，强调了当前存在的一些问题，并实施了改进的反垃圾邮件模型。在这方面，提出了一种基于多自然语言反垃圾邮件（MNLAS）模型的新代理。 MNLAS 模型在电子邮件的垃圾邮件过滤过程中处理视觉信息，如图像和英语和阿拉伯语的文本。 Jade 代理平台和 Java 环境用于实现 MNLAS 模型。 MNLAS 模型在 200 封电子邮件的数据集上进行测试，结果表明它能够高精度地检测和过滤各种垃圾邮件。

**关键词:** 反垃圾邮件分类，机器学习，软件代理，多代理系统。

## I. INTRODUCTION

Spam is the common term for unwanted emails, though more technically spam is defined as Unsolicited Commercial Email (UCE) or Unsolicited Bulk Email (UBE). Here the key distinction be-tween spam and marketing is the unsolicited element i.e. an email “sent to an account by a person unacquainted with the recipient [1].

Spam email is evolving into malicious and provoking additions to internet technology. Keeping an updated blacklist have been made more difficult because of the rapid emergence of distributed phishing attacks and phishing websites. [2] Conventional anti-spam defenses lack the capability to handle the huge volumes of spam types that are beyond their capacity. The increase in spam-related problems makes it important to have tools that are effective and efficient in controlling spam problems. [1]

Researchers have been provided with more appropriate methods of combating spam through machine learning approaches. [3] Anti-spam classification is one of the fields that machine learning is successfully handled. Since spam is contained in the email, spam can be categorized using machine learning approaches like neural network, random forest and support vector machine e.g., [1], [3] and [4]. The emails classification by machine learning techniques to valid or spam email has a lesser human intervention and gives higher accuracy. [5] The classification results depend on the emails content and the suitability of the used techniques.

The email has subject and body data. The following steps are required in order to apply these techniques in the filtration and classification of the emails. The first step is transferring the email contents into a numeric data. The second step is checking and identifying the similarity between the data in the header and the body of the email.

However, it is still possible to have header and body that are not similar and do not reveal the exact information which the sender is trying to convey to the receiver. Thus, the accuracy of the email spam filter can be reduced when headers alone are used. This process is known as preprocessing, and it involves the extraction of the feature, reading and tokenizing, selection of a feature, stemming process and removal of stop words. [2] The pre-processing steps can be a factor in enhancing the efficiency of the text-based spam filtering approach. This step produces feature vectors that are used in training the classifiers and testing the classification outcomes. [3]

The text-based approaches are limited by two main constraints. [2] Firstly, different confusion tricks can be employed in order to affect the identification accuracy of the text-based anti-spam filtering techniques. Secondly, as the internet scale and capacity expand, the kind of information contained in the email is diversified. Moreover, the email evolved text and multimedia objects. [4] Multimedia objects, especially images can be included within the text by legitimate message senders to make the message rich. Image spam such as GIF file format includes an embedded hyperlink that might be linked to a website. [6] The efficiency of the current text-based techniques is greatly reduced by these limitations. A new technique that considers visual content along with the text rep-resents the current research challenge. [7]

In this article, a new Multi-Natural Language Anti-Spam (MNLAS) model which has the capability of overcoming these limitations and cover English and Arabic languages is proposed.

The model is operated by a multi-agent system and made up of visual information, especially images in the anti-spam filtering process. The use of Jade agent platform is employed in implementing the MNLAS within a Java environment. With the use of this application, various kinds of spam emails are detected and filtered successfully.

The remaining part of the paper is prepared into five sections and several subsections. The following section (Section II) reviews the literature on anti-spam classification methods. Section III provides a description of the used research materials and methods. Section IV illustrates the theoretical formulation of the proposed MNLAS model. Subsequently, Section V presents the implementation and the obtained results. Section VI includes the conclusion and some few related future work.

## II. LITERATURE REVIEW

### A. Email

Electronic mail (email) is a messaging system that an electronically transmits messages across computer networks. [8] In this messaging system, a message sender has to open a message panel, includes the recipient's email address, type the message title, then type the message details and send the message by clicking on the send button. It is possible for users to gain access to any free email services like Hotmail, Gmail, Yahoo, or even register with Internet Service Providers (ISPs) to get an email account. This account is free of charge and it only requires an Internet connection. More so, the recipient can receive email almost immediately after sending one. With the use of e-mail, users can communicate with each other at a low cost through an efficient mail delivery system. [9]

Email services have become the most popular and chosen communication tool, because of its availability, reliability and user-friendliness. Therefore, individual users and businesses heavily depend on this communication tool for information and knowledge sharing [10]. As mentioned earlier, email messages consists of header and body of the message. The header is made up of information regarding the transportation and title of the message. The following make up the header information:

a) *From: It holds the details of the sender, such as email address.*
b) *To: It holds the details of the receiver, such as an email address.*
c) *Date: It holds the email sending date to a particular recipient or recipients.*
d) *Received: It holds the intermediary information of the server and the processed date of the email message.*

Using email, which is very fast and cost-effective, the cost of communication can be drastically reduced. More so, the email is a strong tool for marketing which businesses can take advantage of [11]. This technology can be capitalized upon by businesses because it is a widely used advertising tool. However, spam is a problem, which is caused by the simplicity and cost effectiveness of sending emails. Spam is described as bulk unsolicited email that is arbitrary sent to users for different purposes including commercial or phishing reasons. [12].

### B. The Spam Phenomenon

Spam, which is called junk mail is defined in different ways; such definitions explain the difference between spam and legitimate mail (also known as ham, non-spam or genuine mail) Among the most common definitions, the shortest one, is that which defines spam as "unsolicited bulk email". [13], [14], [15]. The following table presents different categories of spam based on Ferris Research. [2]

Table 1. The categories of spam applications

| Categories | Descriptions |
|---|---|
| Health | The spam of fake medications |
| Promotional products | The spam of fake fashion items like clothes bags and watches |
| Adult content | The spam of adult content of pornography and prostitution |
| Finance & marketing | The spam of stock kiting, tax solutions and loan packages |
| Phishing | The spam of phishing or fraud |

| | |
|---|---|
| | such as "Nigerian 419" and "Spanish Prisoner" |
| Malware | The malware or Trojan spam attempts to attack and spy on personal computers |
| Education | The spam of fraudulent certificates of online studies |
| Political | The spam of political targets such as elections by online voting |

Generally, the use of spam is employed in the advertisement of all sorts of services and goods, and the percentage of advertisement devoted to a certain type of services and goods evolves over time. Quite often, the needs of online fraudsters are met by spam. Phishing is a typical case of spamming activity, which involves searching for sensitive information like credit card details and passwords. This search is performed by mimicking official requests from trustworthy authorities, such as service providers or server administration and banks. [16] Viruses are another kind of malicious spam, which sometimes involves the use of a massive spam attack to disrupt the work of a mail server. [17] In summary, a spam message sender aims to; advertise some goods, ideas or services, ideas, use people's private information to cheat them, temporarily crash a mail server, or deliver malicious software. In relation to content, spam is segmented into different subjects and many genres due to the simulation of different classes of legitimate emails like letters, memos and order confirmations. [18] There is a difference between the characteristics of spam traffic and that of legitimate email traffic; specifically, legitimate emails are sent at diurnal periods, whereas spam coming rate is steady over time. [19]

Spammers often use different methods to hide their identity during sending spam, but they usually do not hide their identity during harvesting email addresses from online materials like websites or papers. This implies that the identity of the spammers can be known through the recognition of harvesting activities. [20] One important point to note is that spammers reactive, which means that spammers actively oppose every successful anti-spam effort. [29] This causes the efficiency of every new method to decrease after it is deployed. The evolution of spamming techniques was analyzed by Pu and Webb [21], who revealed that spams construction become ineffective when filters are efficient enough to deal with them, or if other effective measures are taken to combat them.

### C. The Spammers Tricks

For spammers to be able to send spam messages, they must first harvest addresses in order to obtain email addresses; these addresses are harvested from the internet using specialized software. The software is capable of systematically gathering email addresses from websites or discussion groups. [22] In addition, a spammer can purchase or lease sets of email addresses from other specialist co-ops of spammers. In Table 2, the different traps utilized by spammers to stay away from identification via spam filters are explained briefly.

Table 2. Tricks used by spammers to send spam [2]

| Tricks | Descriptions |
|---|---|
| Zombies or Botnets | This involves sending off a huge number of viruses, spam and malware by PCs on the Internet. |
| Bayesian sneaking and poisoning | This focuses on using words that are not commonly used in spam messages to write a spam message. More so, the Bayesian filter database is not "poisoned" by the spam messages. |
| IP address | Acquiring or utilizing an IP address that has a trusted or nonpartisan reputation. |
| Offshore ISPs | The use of offshore ISPs that lack security measures. |
| Open proxies / open-relay servers | Focuses on using servers for the re-direction of spam to vulnerable users. |
| Third-party mail back software | Utilizing inappropriately anchored email uses on honest sites. |
| Falsified header information | The addition of fake header information to the spam message. |
| Obfuscation | Splits messages or words through the use of nonsense creative symbols of HTML tags in order to disguise spam messages. |
| Vertical slicing | Involves the vertical writing of spam messages. |
| HTML | Aims at avoiding detection of spam message |

| | |
|---|---|
| manipulation | by manipulating HTML format. |
| HTML encoding | Converts binary attachment into plain text characters by using an encoding technique like Base64. |
| JavaScript messages | Here entire substance of the spam messages is set in a JavaScript scrap, which is enacted once the message is opened. |
| ASCII art | Writing spam messages using letter glyphs of standard letters. |
| Image-based | Textual information sends by using the image. |
| URL address or redirect URL | Here the addition of URL address is made in order to avoid detection. It sometimes involves the use of irrelevant "portals" to lead the user to their actual websites. |
| Encrypted messages | Encoding message where it just unscrambled once it achieves the letterbox. |

### D. Spam Impact

The first email to be considered spam was sent on May 3rd, 1978 to about 400 ARPANET users and it was an open invitation to upcoming computer hardware demonstrations1. [23] Since then they have become a daily nuisance for the end users of any email service. The current volume of global email traffic that is spam is estimated to be in the region of 53%2 – 58%3 down from a peak of 88% in 2010 [24] though still well above its estimated proportion of 10% in 1998. [25] Figure 1 shows the average of spam distribution from 2006 to 2016.

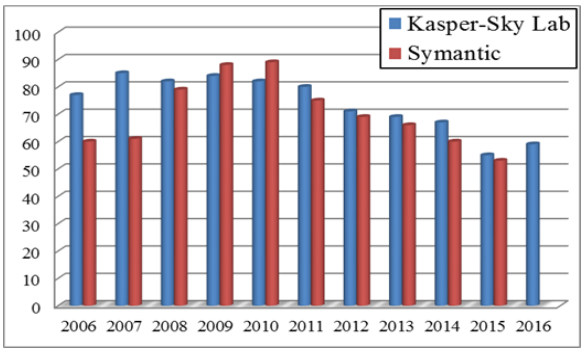


Figure 1.Average of spam from 2006 to 2016 [2], [9]

### E. Existing Solutions for Spam

In this section different anti-spam methods are presented. Some of them are listed in Table 3. The methods use spam filters that can prevent spam emails from getting to their target and causing any harm. The contributions and methods are highlighted in the review, and some weaknesses of the approaches are identified.

Table 3.
Anti-spam methods

| Author | Descriptions |
|---|---|
| Sankar et al. [8] | In this method, masked spam is detected using synonym relation completion and keyword concatenation. The content of the message is used in predicting the spam category, instead of depending on a set of predetermined keywords. |
| Bhowmick et al. [9] | In this study, the authors basically focused on Machine Learning-based spam filters as well as their variations. Based on their review of a wide range of ideas related to email spam, a report on efforts made to address the problems of spamming, effectiveness of existing techniques of spam filtering and current progress was generated. |
| Mirza et al. [10] | This study provides an explanation of the different categories of emails that can be used in differentiating between spam emails and non-spam emails. In order to achieve this, the Naive Bayesian Classifier was used in creating an email classification system to classify spam and not spam. |
| McGetrick et al. [11] | Here, personality insight and language tone scores were created using the text from emails; the use of IBM Watsons' Tone Analyzer API was employed. The use of those scores is employed in investigating whether the language used in emails can be changed into useful features, which can be utilized incorrectly classifying them as spam or genuine emails. Furthermore, machine learning techniques are used in the course of the investigation. |
| Bassiouni et al. [12] | This group of researchers focused on the classification of spam emails from Inboxes. The application of ten different classifiers is made to one benchmark dataset for evaluating the performance of each classifier, to determine which of them produces better results. |

Usually, there are numerous methods that are accessible to spam, for example, utilizing sender area checking, contents checking, open transfer master habitation and checking the IP address or space names. [26] Therefore, spammers effectively conquer these basic forms with progressively refined variations of spam to avoid identification. The measures connected with to control spam are; Legislation methods, Black-list and white-list, Heuristic approaches, Machine learning approaches, Naïve Bayesian

Classification, Support Vector Machine Classification, Neural Network Classification. [27]

A mobile agent system was proposed by Mahmoud et al. [29] for the purpose of identifying and blocking SMS spam messages. In their work, they tried to provide protection for smartphone by filtering SMS spam that contains abbreviations and idioms. The system functions based on Artificial Immune System (AIS) and Naïve Bayesian (NB) algorithm. A set of features are produced by the AIS. With the use of the NB algorithm, the messages are classified based on the extracted features. This system is tested by a SMS messages dataset of 1324 messages. Results of the classification reveal that, the false positive rate is 6%, the detection rate is 82% and the overall accuracy is 91% on average.

A random forest algorithm is proposed by Akinyelu and Adewumi [1] as a machine learning classifier in order to identify phishing emails. The random forest algorithm is employed in classifying 2000 emails. The objectives of the research are to reduce the feature and increase the classification accuracy. A high level of classification accuracy of up to 99.7% with a little amount of 0.06% false positive is achieved by the proposed algorithm. Nevertheless, the research scope only covered the classification aspect without considering vital information which can affect the results, especially, in the case of limited text in the email.

Yüksel et al. [3] attempted to resolve the problem of spam by hindering the spam emails from being distributed within the email systems. To achieve this, they propose a cloud base system, which involves the identification of spam emails using analytics and machine learning algorithms. The algorithms are support vector machine and decision tree. The system is tested by Azure testing platform and the results of the tests show that the support vector machine attains the higher accuracy of up to 97.6% and false positive rate of 2.33%. The decision tree attains the lower accuracy of 82.6% and false positive rate of 17.3%. The results further reveal that the increase in the spam volumes is influenced by the number of received emails. An optimal spam detection technique was proposed by Lee et al. [29].

## III. RESEARCH METHODOLOGY

This study incorporates the fundamental research stage, the data gathering stage and mutual document processing stage. These three stages depict the study techniques for the research.

### A. The Phase of Preliminary Study

This work stage aims at making an anti-spam method based on joint document processing of emails data utilizing agent technology and machine learning techniques. Based on the earlier mentioned methods, in this phase, the researcher reveals the extant anti-spam approaches. The results specify the requirements for the construction of an enhanced visual anti-spam scheme.

### B. The Phase of Data Collection

The focus of this stage is to set up the testing dataset. The stage includes associating an email to the outlook to have the capacity to peruse messages from the viewpoint. Afterwards, the emails are stored in the form of HTML, and WordPad is used in saving the emails as text files. This out-come in the substance of the email with their HTML body. These two arrangements of messages are gathered as spam messages, and real messages. Each of them contains 100 emails, i.e., 200 emails in total. Subsequently, testing is carried out using these emails.

### C. The Phase of Mutual Document Processing

In this phase, the email is changed to a uniform format, which the learning algorithm can under-stand and process. [5], [6] The email data is split into a header and body parts. The header part contains the general information of the sender like the subject, recipient's details and route details are included. The body part contains the message main content. Prior to running the process of filtration, the extraction of this information must be performed through

preprocessing. The header and the body can be different, and they do not point out the exact information that the sender wants to convey to the receiver. Therefore, the accuracy of the email spam filtering can be reduced if just one part of the email is used (either header or body).

## IV. The Proposed Model

The operation of the MNLAS model is based on a multi-agent system which interact with each other in the anti-spam classification environment [30], [31]. These agents do not operate independently, rather the work together to solve problems that cannot be independently solved [32], [33]. The flexibility of the system is increased using the agent application [34]. The agent application is also capable of segregating the functionality of the system [14] while enabling interaction between the system and its modules [35], [36]. The MNLAS model is made up of five agents as shown in Figure 2.

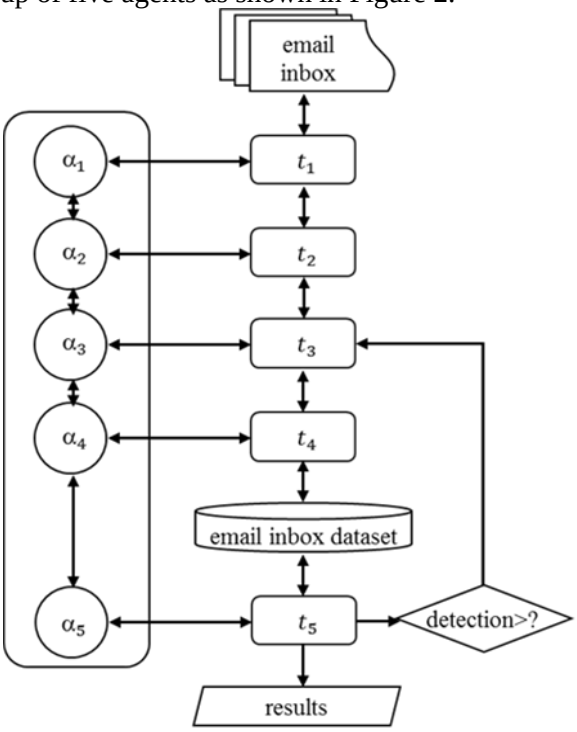


Figure 2.The MNLAS model

Figure 2 illustrates the processing steps of the MNLAS model by the multi-agent system. All these agents carry out specific tasks in the process of spam detection as follows: (1) $\alpha_1$: It represents the first agent that performs processing the short words task, $t_1$, (2) $\alpha_2$: It represents the second agent that performs the features extraction task, $t_2$, (3) $\alpha_3$: It represents the third agent that performs the features selection task, $t_3$, (4) $\alpha_4$: It represents the fourth agent that performs the instances presentation task, $t_4$ and (5) $\alpha_5$: It represents the fifth agent that performs the email classification task, $t_5$.

### A. The Agent for Short Words Form

Usually, spam emails start with short forms to confuse the spam engines. According to Fung [37], software for translating Short Message Service (SMS) have been developed by a few companies [38]. In this work translation, the researchers consider using the original long form of words to replace short words so that the anti-spam engine will not be confused [6]. For instance, the word ROFL, which means “rolling on the floor laughing” is meaningless for the spam engine. For this reason, 1300 abbreviations were published by Beal [37]. Some of the abbreviations are presented in Table 4.

Table 4.
Example of short messages forms

| Abb. | Meaning | Abb. | Meaning |
|---|---|---|---|
| ? | I have a question | 4 | Short for "for" in SMS |
| @TEOTD | At the end of the day | 4EAE | Forever and ever |
| ?4U | I have a question for you | 404 | I don't know |
| 2 | Meaning "to" in SMS | 86 | Over |
| ^^ | Meaning "read line" or "message above" | 420 | Lets get high |
| 10X | Thanks | 6Y | Sexy |

The anti-spam efforts are challenged by the existence of the short form of words; this is a new challenge. In order to identify abbreviations or short words, a table of 1300 short words obtained from [37] is used the agent of the MNLAS model.

### B. The Agent for Features Extraction

This agent extracts features from email header and body to for the vector space. The text message is reduced to its colloquial component by the agent using a tokenization method [39]. The message is taken and split into different tokens (words). By making the addition of the words to the vector space and constructing a features space for classification, these words are parsed [8]. In the process of tokenization, all the

features are extracted from the message without regarding its relevance. Tokenized features is a technique that might be highly vulnerable to confusing of contents. The tokenization of features involves reduction of dimension, stemming and the processor removing stop-word [40], [41].

This work takes in to account the possibility of Multi-Natural Languages (MNL) emails that include English words characters and/or numbers in written Arabic messages. Some such examples are contained in Table 5.

Table 5.
An example of Arabic-English letters

| Character in Arabic | Character in English | Character in Arabic | Character in English |
|---|---|---|---|
| ئ,ء,أ | 2 | ب | b |
| غ | ‘3 | ف | f |
| ج | G | ح | 7 |
| ش | Sh | ص | 9 |
| غ | ‘3 | ف | f |
| ز | Z | س | s |

The use of English words when chatting in Arabic has become popular, and the use of this technique was initiated by spammers. The use of the same technique can also be employed in the English language and other languages. Table 6 presents an example of a modern Arabic chatting language.

Table 6.
English-Arabic alternatives

| English | Arabic | Alternative |
|---|---|---|
| The student went home” | ذهب الطالب الى البيت | “4ahaba al6aleb 2ela albeit” |

In order to provide a flexible and brief means of text strings matching, regular expressions \such as a character, characters and patterns of characters or words are used by the agent. The purpose of using regular expressions is to examine MNL text. Few specifications of the regular expressions are given in the examples below: (1) The characters "car" in any context, such as "car", "career ", or "carrageen ", (2) The word "car" when it appears as an isolated word, (3) The word "car" when preceded by the word "blue" or "red" and (4) A dollar signs immediately followed by one or more digits, and then optionally a period and exactly two more digits (for example "$10", or "$245.99").

### C. The Agent for Features Selection

The dimensions of the features vector are reduced by the selection agent [8], which measures the frequency of the occurrence of a certain word or phrases. Based on the predetermined threshold, words and phrases that are irrelevant and idle are eliminated. This process is crucial to the improvement of the process of classification [2]. In this process, phrase with mutual information, commoner morphological and in flexional endings and the Stop Word (e.g., “a”, “an” and “the”) are eliminated from English words. A sample of stop words is presented in Table 7.

Table 7.
A sample of stop words

| | | | | | |
|---|---|---|---|---|---|
| a | did | herself | not | the | we've |
| about | didn't | him | of | their | were |
| above | do | himself | off | theirs | weren't |
| after | does | his | on | them | what |
| again | doesn't | how | once | themselves | what's |
| against | doing | how's | only | then | when |

### D. The Agent for Instances Representation

The function of the agent is to represent the features into the right format for the classification step. The use of n-gram technique is employed in representing the features in numerical and lexical and numerical forms. A vector of features which is brief and organized is generated by the n-gram; the summary of the email text is contained in this vector of features [42].

Let $x$ equals a number of words in a given feature of a feature vector $v_{i,j}$, the number of n-grams for the vector can be found by:

$$nGrams(v_i) = x_i - (n-1) \quad (1)$$

### E. The Agent for Classification

In the MNLAS model, the use of a random forest classification algorithm, which is adapted from [1] is employed by the classification agent; the algorithm is applied with three different training and testing phases. During the training period, the sets of feature vectors, which are set in the previous phase to train the classifier, are used by the agent [43], [44], [45]. The training includes 3-fold cross-validation data allocations. The classifier is run by the agent in the testing

phase to differentiate the spam emails, $E^s$, from the legitimate emails, $E^l$.

## V. TESTING AND RESULTS

The MNLAS model is a machine learning approach that employs a multi-agent system to operate the identification of spam emails. The implementation of the MNLAS model is performed using a Jade agent platform within a Java environment. The application can detect and filtering spam from non-spam emails of various kinds successfully. The ratio between the number of legitimate emails and correctly classified spam emails to the total number of emails used in the testing phase represents the accuracy of the identification.

$$Accuracy = (T^p + T^n)/(T^p + E^s + F^p + T^n) \quad (2)$$

where the $F^n$ denotes false negatives that results from the $E^s$ that are classified as $E^l$; the $F^p$ denotes false positives that results from the $E^l$ that are classified as $E^s$; $T^p$ denotes true positive that results from the $E^s$ that are correctly identified; $T^n$ denotes true negative that results from the $E^l$ that are correctly identified;

Table 8 shows the classification results of the performed test. The results of the high accuracy of the MNLAS model are contained in the three rungs of Table V. The model can differentiate legitimate emails from spam with an average accuracy of 91.90%. This research outcome support the application of this work in real-life situations.

Table 8.
The classification results

| Run | Allocation | Precision | Recall | F-measure | Accuracy(%) |
|---|---|---|---|---|---|
| 1 | 60-40 | 0.9026 | 0.8488 | 0.8949 | 89.86% |
| 2 | 50-50 | 0.9370 | 0.9118 | 0.9242 | 92.52% |
| 3 | 40-60 | 0.9026 | 0.9180 | 0.9321 | 93.32% |

## VI. CONCLUSION AND FUTURE WORK

Currently, the use of spam emails in phishing has become a huge challenge to computer security. In this regard, various methods of machine learning are used to identify anti-spam approaches; some of the approaches include decision tree, random forest and support vector machine. Suitable solutions have been provided by these methods, but these methods also have their impediments that outcome from and are identified with the spam email substance just as conditions like the presence of short messages, MNL messages and pictures. The greater part of the works around there utilized instant datasets, which reject these restrictions. The main point of related studies is on enhancing the precision of classifier. In this study, an endeavor is made to proffer outcomes by proposing the Multi-Natural Language Anti-Spam (MNLAS) scheme. In the task of filtering spam, the MNLAS uses short messages, visual information and texts of an email. The model can separate spam from real messages with a normal exactness of 91.90%. The outcomes bolster the utilization of this work in genuine situations. The future work ponders testing the model with various standard datasets.

**Acknowledgment**

This project is sponsored by the Universiti Tun Hussein Onn Malaysia (UTHM) under Tier 1 research grant scheme Vot U891.

## REFERENCES

[1] AKINYELU, A. A., & ADEWUMI, A. O. (2014). Classification of phishing email using random forest machine learning technique. Journal of Applied Mathematics.

[2] SUBRAMANIAN, T., TAQA, A.Y., JALAB, H.A, (2010). Overview of textual anti-spam filtering techniques. Int. J. Phys. Sci. 5 (12:1869-1882.

[3] YÜKSEL, A. S., CANKAYA, S. F., & ÜNCÜ, İ. S. (2017). Design of a Machine Learning Based Predictive Analytics System for Spam Problem. Acta Physica Polonica, A., 132(3).

[4] MOHAMMED,M. A., GUNASEKARAN, S. S., MOSTAFA, S. A., MUSTAFA, A., & GHANI, M. K. A. (2018, August). Implementing an Agent-based Multi-Natural Language Anti-Spam Model. In 2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR) (pp. 1-5). IEEE.

[5] NAZIROVA, S. (2011). Survey on spam filtering techniques. Communications and Network, 3(03), 153.

[6] CARUANA, G., & LI, M. (2012). A survey of emerging approaches to spam filtering. ACM Computing Surveys (CSUR), 44(2), 9.

[7] SHAFI'I, M. A., LATIFF, M. S. A., CHIROMA, H., OSHO, O., ABDUL-SALAAM, G., Abubakar, A. I., & Herawan, T. (2017). A review on mobile SMS spam filtering techniques. IEEE Access, 5, 15650-15666.

[8] SANKAR, K. V., UMA, S., SUBIN, P. S., & ABHIMANNAN, T. (2015). Mask Spam Detection Using Difficult Key-word Identification and Relation Completion.

[9] BHOWMICK, A., & HAZARIKA, S. M. (2016). Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends. arXiv preprint arXiv:1606.01042.

[10] MIRZA, N., PATIL, B., MIRZA, T., & AUTI, R. (2017, June). Evaluating efficiency of classifier for email spam detec-tor using hybrid feature selection approaches. In Intelligent Computing and Control Systems (ICICCS), 2017 International Conference on (pp. 735-740). IEEE.

[11] MCGETRICK, C. (2017). Investigation into the Application of Personality Insights and Language Tone Analysis in Spam Classification.

[12] BASSIOUNI, M., Ali, M., & EL-DAHSHAN, E. A. (2018). Ham and Spam E-Mails Classification Using Machine Learning Techniques. Journal of Applied Security Research, 13(3), 315-331.

[13] FERRARA, E. (2018). Measuring social spam and the effect of bots on information diffusion in social media. In Complex Spreading Phenomena in Social Systems (pp. 229-255). Springer, Cham.

[14] BOUDIA, M. A., RAHMANI, M. E., & RAHMANI, A. (2018). Comparative Study between a Swarm Intelligence for Detection and Filtering of SPAM: Social Bees vs. Inspiration From the Human Renal. In Handbook of Re-search on Biomimicry in Information Retrieval and Knowledge Management (pp. 38-65). IGI Global.

[15] PENG, W., HUANG, L., JIA, J., & INGRAM, E. (2018, August). Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection. In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 849-854). IEEE.

[16] SINGH, M. (2019). Classification of Spam Email Using Intelligent Water Drops Algorithm with Naïve Bayes Classifier. In Progress in advanced computing & Intelligent Engineering (pp. 133-138). Springer, Singapore.

[17] BLANZIERI, E., & BRYL, A. (2008). A survey of learning-based techniques of email spam filtering. Artificial Intel-ligence Review, 29(1), 63-92.

Guerra, P. H. C., Guedes, D., Wagner Meira, J., HOEPERS, C., CHAVES, M. H. P. C., &

[18] STEDING-JESSEN, K. (2010). Exploring the spam arms race to characterize spam evolution. In Proceedings of the 7th Collabor19.

[19] Ation, Electronic messaging, Anti-Abuse and Spam Conference (CEAS), Redmond, WA.

[20] SKUDLARK, A. E., ERMAN, J., JIN, Y., SRIVASTAVA, A., & TRAN, L. K. (2017). U.S. Patent No. 9,584,989. Washing-ton, DC: U.S. Patent and Trademark Office.

[21] TU, H., DOUPÉ, A., ZHAO, Z., & AHN, G. J. (2016, May). Sok: Everyone hates robocalls: A survey of tech-niques against telephone spam. In Security and Privacy (SP), 2016 IEEE Symposium on (pp. 320-338).

[22] CHEN, L., YAN, Z., ZHANG, W., & KANTOLA, R. (2015). TruSMS: a trustworthy SMS spam control system based

on trust management. Future Generation Computer Systems, 49, 77-93.
[23] Ogudo, K.A.; Muwawa Jean Nestor, D.; Ibrahim Khalaf, O.; Daei Kasmaei, H. A Device Performance and Data Analytics Concept for Smartphones' IoT Services and Machine-Type Communication in Cellular Networks. *Symmetry* **2019**, *11*, 593
[24] FAWCETT, T. (2003). In vivo spam filtering: a challenge problem for KDD. ACM SIGKDD Explorations News-letter, 5(2), 140-148.
[25] RAAD, M., YEASSEN, N. M., ALAM, G. M., ZAIDAN, B. B., & ZAIDAN, A. A. (2010). Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing. African Journal of Business Management, 4(11), 2362-2367.
[26] GOODMAN, J. (2004, July). IP Addresses in Email Clients. In CEAS.
[27] GUPTA, H., JAMAL, M. S., MADISETTY, S., & DESARKAR, M. S. (2018, January). A framework for real-time spam detection in Twitter. In Communication Systems & Networks (COMSNETS), 2018 10th International Confer-ence on (pp. 380-383).
[28] ADEWOLE, K. S., ANUAR, N. B., KAMSIN, A., & SANGAIAH, A. K. (2017). SMSAD: a framework for spam mes-sage and spam account detection. Multimedia Tools and Applications, 1-36.
[29] MAHMOUD, T. M., & MAHFOUZ, A. M. (2012). SMS spam filtering technique based on artificial immune system. International Journal of Computer Science Issues (IJCSI), 9(2), 589.
[30] LEE, S. M., KIM, D. S., KIM, J. H., & PARK, J. S. (2010, February). Spam detection using feature selection and parameters optimization. In Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on (pp. 883-888). IEEE.
[31] MOSTAFA, S. A., MUSTAPHA, A., HAZEEM, A. A., KHALEEFAH, S. H., & MOHAMMED, M. A. (2018). An Agent-Based Inference Engine for Efficient and Reliable Automated Car Failure Diagnosis Assistance. IEEE Access, 6, 8322-8331.
[32] JASSIM, O. A., MAHMOUD, M. A., & AHMAD, M. S. (2015). A multi-agent framework for research supervision management. In Distributed Computing and Artificial Intelligence, 12th International Conference (pp. 129-136). Springer, Cham.
[33] MOSTAFA, S. A., MUSTAPHA, A., MOHAMMED, M. A., AHMAD, M. S., & MAHMOUD, M. A. (2018). A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application. International journal of medical informatics, 112, 173-184.
[34] MAHMOUD, M. A., AHMAD, M. S., AHMAD, A., YUSOFF, M. Z. M., MUSTAPHA, A., & HAMID, N. H. A. (2013, May). Obligation and Prohibition Norms Mining Algorithm for Normative Multi-agent Systems. In KES-AMSTA (pp. 115-124).
[35] Benjamin Durakovic, et al, "Lean Manufacturing: Trends and Implementation Issues", Periodical of Engineering and Natural Sciences, Vol. 6, No. 1, pp. 130-143 (2018), ISSN: 2303-4521.
[36] MAHMOUD, M. A., AHMAD, M. S., AHMAD, A., YUSOFF, M. Z. M., & MUSTAPHA, A. (2012). Norms detection and assimilation in multi-agent systems: a conceptual approach. In Knowledge Technology (pp. 226-233). Springer, Berlin, Heidelberg.
[37] MOSTAFA, S. A., AHMAD, M. S., MUSTAPHA, A., & MOHAMMED, M. A. (2017). Formulating layered adjustable autonomy for unmanned aerial vehicles. International Journal of Intelligent Computing and Cybernetics, 10(4), 430-450.
[38] FUNG L.M (2005), "SMS Short Form Identification and Codec", Unpublished master's thesis, National Univer-sity of Singapore, Singapore. No. of pages (32).
[39] BEAL V. (2010) Text Messaging and Chat Abbreviations: A Guide to Understanding Text Messages, Chat Ab-breviations, and Twitter Messages.

[40] ZDZIARSKI JA (2005). Tokenization: The Building Blocks of Spam. In Pollock W, ZINKANN E (Eds.), Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification San Francisco: No Starch Press, pp. 97-110.
[41] GUZELLA TS, CAMINHAS WM (2009). A review of machine learning approaches to spam filtering. Expert system with Application, 36: 10206-10222.
[42] PAGANI, F., De ASTIS, M., GRAZIANO, M., LANZI, A., & BALZAROTTI, D. (2016, June). Measuring the Role of Grey-listing and Nolisting in Fighting Spam. In Dependable Systems and Networks (DSN), 2016 46th Annual IEEE/IFIP International Conference on (pp. 562-571).
[43] MAHMOUD, M. A., & AHMAD, M. S. (2016, August). A prototype for context identification of scientific papers via agent-based text mining. In 2016 2nd International Symposium on Agent, Multi-Agent Systems and Robot-ics (ISAMSR) (pp. 40-44). IEEE.
[44] MAHMOUD, M. A., AHMAD, M. S., & Yusoff, M. Z. M. (2016). A conceptual automated negotiation model for decision making in the construction domain. In Distributed Computing and Artificial Intelligence, 13th Interna-tional Conference (pp. 13-21). Springer, Cham.
[45] KHALAF, B. A., MOSTAFA, S. A., MUSTAPHA, A., MOHAMMED, M. A., ABDUALLAH, W. M. (2019). Comprehensive Review of Artificial Intelligence and Statistical Approaches in Distributed Denial of Service Attack and Defense Methods. IEEE Access, 7, 51691-51713.

## 参考文:

[1] AKINYELU, A. A., & ADEWUMI, A. O. (2014). 使用随机森林机器学习技术分类钓鱼邮件 应用数学杂志
[2] SUBRAMANIAN，T.，TAQA，A.Y.，JALAB，H.A，（2010）文本反垃圾邮件过滤技术概述 诠释 J. Phys 科学 5 (12): 1869-1882.
[3]YÜKSEL，A.S.，CANKAYA，S。F.，& ÜNCÜ，İ. S.（2017）基于机器学习的垃圾邮件问题预测分析系统设计 Acta Physica Polonica，A.，132 (3)
[4] MOHAMMED，M. A.，GUNASEKARAN，S.S.，MOSTAFA，S.A.，MUSTAFA，A.&GHANI，M.K.A.(2018，August) 实现基于代理的多自然语言反垃圾邮件模型 2018 年代理，多代理系统和机器人国际研讨会 (ISAMSR) (第 1-5 页) IEEE
[5] NAZIROVA，S.（2011 年）垃圾邮件过滤技术调查 通讯与网络，3(03)，153
[6] CARUANA，G.，&LI，M.(2012) 对垃圾邮件过滤新兴方法的调查。 ACM 计算调查 (CSUR)，44 (2)，9.
[7] SHAFI'I，M。A.，LATIFF，M。S. A.，CHIROMA，H.，OSHO，O.，ABDUL-SALAAM，G.，Abubakar，A.I.，& Herawan，T.(2017) 移动短信垃圾邮件过滤技术综述 IEEE Access, 5, 15650-15666
[8] SANKAR，K.V.，UMA，S.，SUBIN，P.S.，& ABHIMANNAN，T.（2015）使用困难的关键字识别和关系完成的面具垃圾邮件检测
[9] BHOWMICK，A.，& HAZARIKA，S。M.（2016）电子邮件垃圾邮件过滤的机器学习: 回顾，技术和趋势 arXiv preprint arXiv：1606.01042
[10] MIRZA，N.，PATIL，B.，MIRZA，T.，& AUTI，R.（2017 年 6 月）使用混合特征选择方法评估电子邮件垃圾邮件检测器的分类器效率 在智能计算和控制系统 (ICICCS)，2017 年国际会议（第 735-740 页） IEEE。
[11] MCGETRICK, C. (2017). Investigation into the Application of Personality Insights and Language Tone Analysis in Spam Classification.
[12] BASSIOUNI，M.，Ali，M。，&EL-DAHSHAN，E.A。（2018）使用机器学习技术的火腿和垃圾邮件电子邮件分类 Journal of Applied Security Research，13（3），315-331
[13] FERRARA，E.（2018 年）衡量社交垃圾邮件和机器人对社交媒体中信息传播的影响在社会系统中的复杂传播现象（第 229-255 页）施普林格，湛
[14] BOUDIA，M.A.，RAHMANI，M。E.，&RAHMANI，A.（2018）群体智能检测与 SPAM 过滤的比较研究：社会蜜蜂与人类肾

脏的启示。“信息检索和知识管理中的生物模拟研究手册”（第 38-65 页） IGI Global
[15] PENG，W.，HUANG，L.，JIA，J.，& INGRAM，E.（2018，August）通过智能文本修改检测增强朴素贝叶斯垃圾邮件过滤器 2018 年第 17 届 IEEE 国际计算与通信信任，安全与隐私会议/第 12 届 IEEE 大数据科学与工程国际会议 (TrustCom / BigDataSE) (第 849-854 页) IEEE
[16] SINGH，M. (2019 年) 基于 NaïveBayes 分类器的智能水滴算法对垃圾邮件的分类 高级计算与智能工程进展 (第 133-138 页) 施普林格，新加坡
[17] BLANZIERI，E，& BRYL，A. (2008) 基于学习的电子邮件垃圾邮件过滤技术调查人工智能评论，29（1），63-92 Guerra, P. H. C., Guedes, D., Wagner Meira, J., HOEPERS, C., CHAVES, M. H. P. C., &
[18] STEDING-JESSEN，K. (2010) 探索垃圾邮件军备竞赛以描述垃圾邮件的演变特征在第 7 期 Collabor19 的会议录中
[19] Ation，电子信息，反滥用和垃圾邮件会议 (CEAS)， 华盛顿州雷蒙德市
[20] SKUDLARK，A.E.，ERMAN，J.，JIN，Y.，SRIVASTAVA，A.，& TRAN，L.K. (2017) 美国专利 No.9,584,989 华盛顿特区: 美国专利商标局
[21] TU，H.，DOUPÉ，A.，ZHAO，Z.，& AHN，G.J. (2016，May) Sok：每个人都讨厌 robocalls：针对电话垃圾邮件的技术调查。在安全和隐私（SP），2016 IEEE Symposium on（pp.320-338）
[22] CHEN，L.，YAN，Z.，ZHANG，W.，&KANTOLA，R.（2015） TruSMS：基于信任管理的值得信赖的 SMS 垃圾邮件控制系统。 Future Generation Computer Systems，49,77-93
[23] Ogudo，K.A.; Muwawa Jean Nestor，D; Ibrahim Khalaf，O.; Daei Kasmaei，H. 智能手机在蜂窝网络中的物联网服务和机器类通信的设备性能和数据分析概念 Symmetry 2019,11,593
[24] FAWCETT，T.（2003）体内垃圾邮件过滤：KDD 的挑战问题。 ACM SIGKDD Explorations News-letter，5（2），140-148
[25] RAAD，M.，YEASSEN，N.M.，ALAM, G.M.，ZAIDAN，B.B.，& ZAIDAN A.A. (2010) 垃圾邮件广告通过电子邮件的影响：一项评估反垃圾邮件对电子邮件营销的影响的研究。非洲商业管理杂志，4 (11)，2362-2367
[26] GOODMAN，J. (2004 年 7 月) 电子邮件客户端中的 IP 地址 在 CEAS
[27] GUPTA，H.，JAMAL，M.S.，MADISETTY，S.，& DESARKAR，M.S. (2018，January) Twitter 中实时垃圾邮件检测的框架 通信系统与网络 (COMSNETS), 2018 年第 10 届国际会议 (第 380-383 页)
[28] ADEWOLE，K.S.，ANUAR，N.B.，KAMSIN，A.，& SANGAIAH，A.K. (2017) SMSAD：垃圾邮件和垃圾邮件帐户检测的框架。多媒体工具和应用程序，1-36
[29] MAHMOUD，T.M.，& MAHFOUZ，A.M. (2012) 基于人工免疫系统的短信垃圾邮件过滤技术。国际计算机科学期刊 (IJCSI)，9(2)，589
[30] LEE，S.M.，KIM，D.S.，KIM，J.H.，&PARK，J.S. (2010，February) 使用特征选择和参数优化的垃圾邮件检测 在复杂，智能和软件密集系统 (CISIS)，2010 年国际会议 第 883-888 页 IEEE
[31] MOSTAFA，S.A.，MUSTAPHA，A.，HAZEEM，A.A.，KHALEEFAH，S.H.，& MOHAMMED，M.A. (2018) 基于代理的推理引擎，用于高效可靠的自动汽车故障诊断辅助 IEEE Access，6,8322-8331
[32] JASSIM，O.A.，MAHMOUD，M.A.，&AHMAD，M.S. (2015) 用于研究监督管理的多代理框架 在分布式计算和人工智能，第 12 届国际会议 (第 129-136 页).施普林格，湛
[33] MOSTAFA，S.A.，MUSTAPHA，A.，MOHAMMED，M.A.，AHMAD，M.S.，& MAHMOUD，M.A. (2018)一种用于自动化老年人运动监测应用的多智能体系统可调自治的模糊逻辑控制 国际医学信息学杂志 112,173-184
[34] MAHMOUD，M.A.，AHMAD，M.S.，AHMAD，A.，YUSOFF，M.Z. M.，MUSTAPHA，A.，& HAMID，N.H. A. (2013, May) 规范多智能体系统的义务和禁止规范挖掘算法 在 KES-AMSTA 第 115-124 页
[35] Benjamin Durakovic 等人，“精益生产: 趋势与实施问题”，工程与自然科学期刊，Vol

6，No.1，pp.130-143 (2018)，ISSN：2303-4521
[36] MAHMOUD，M。A.，AHMAD，M.S.，AHMAD，A.，YUSOFF，M.Z. M.，& MUSTAPHA，A. (2012) 多智能体系统中的规范检测和同化：概念方法 在知识技术 (第 226-233 页) 中 施普林格，柏林， 海德堡
[37] MOSTAFA，S.A.，AHMAD，M.S.，MUSTAPHA，A.，& MOHAMMED，M.A. (2017) 为无人机制定分层可调自治 国际智能计算与控制论杂志，10(4)，430-450
[38] FUNG L.M (2005)，“SMS Short Form Identification and Codec”，未发表的硕士论文 新加坡国立大学 页数 (32)
[39] BEAL V. (2010) 短信和聊天缩写：理解短信，聊天缩写和 Twitter 消息的指南
[40] ZDZIARSKI JA (2005 年) 标记化: 垃圾邮件的构建块 在 Pollock W，ZINKANN E (编辑)，结束垃圾邮件: 贝叶斯内容过滤和统计语言分类艺术旧金山: 没有淀粉出版社，第 97-110 页
[41] GUZELLA T.S.，CAMINHAS WM (2009 年) 回顾垃圾邮件过滤的机器学习方法 应用专家系统，36: 10206-10222
[42] PAGANI，F.，De ASTIS，M.，GRAZIANO，M.，LANZI，A.，& BALZAROTTI，D. (2016 年 6 月) 衡量灰色上市和不存在在打击垃圾邮件中的作用在可靠的系统和网络 (DSN)，2016 年第 46 届 IEEE / IFIP 国际会议 (第 562-571 页)
[43] MAHMOUD，M.A.， 和 AHMAD，M.S. (2016 年 8 月) 通过基于代理的文本挖掘进行科技论文上下文识别的原型 2016 年第二届代理, 多智能体系统和机器人（ISAMSR）国际研讨会 (第 40-44 页) IEEE
[44] MAHMOUD，M.A.，AHMAD，M.S.，& Yusoff，M.Z. M. (2016) 构造领域决策的概念自动协商模型 在分布式计算和人工智能，第 13 届国际会议 (第 13-21 页) 施普林格，湛
[45] KHALAF，B.A.，MOSTAFA，S.A.，MUSTAPHA，A.，MOHAMMED，M.A.，ABDUALLAH，W.M. (2019) 分布式拒绝服务攻防方法中人工智能与统计方法综述 IEEE Access，7, 51691-517