Spam Email Filtering Using Neural Networks

Parikshith T.
Department of Computer Science & Engineering,
Sai Vidya Institute of Technology,
Bengaluru, India.
parikshitht.16cs@saividya.ac.in

Yashas N.
Department of Computer Science & Engineering,
Sai Vidya Institute of Technology,
Bengaluru, India.
yashasn.16cs@saividya.ac.in

Puneerth P.
Department of Computer Science & Engineering,
Sai Vidya Institute of Technology,
Bengaluru, India.
puneethp.16cs@saividya.ac.in

Vignesh N.P.
Department of Computer Science & Engineering,
Sai Vidya Institute of Technology,
Bengaluru, India.
vigneshnp@saividya.ac.in

Dr. Archana R. A.
Department of Computer Science & Engineering,
Sai Vidya Institute of Technology,
Bengaluru, India.
archana.ra@saividya.ac.in

***Abstract*—One of the most secure mediums of communication for the exchange of information on the web is e-mail. An incontrollable surge in admiration, the number of unsolicited data has also surged rapidly. To filter data, diverse approaches are in existence, they implicitly distinguish and confiscate these shaky mails. There are several numbers of email spam filtering techniques such as Recognition-based, congregating, Empirical-based, and so on. In this paper we show the prominent existing methods for classification of spam emails or messages, then based on experimentation of accuracies, we show why the neural network model would be the best choice for building a spam-filtering system. The internet has undeniably become the linking tool for bringing together regulars and business people, nations and territories, continents, and islands regardless of their economic, political, cultural, and social affiliations.**

I. Introduction

In recent eons, the internet has shaped numerous podia for making human life become extra protected. Amongst them, e-mail is an extensive stage for consumer communiqué. Electronic mail (email) is a nonentity, or in lay man's terms, it is an electronic messaging structure that sends information or messages from one user to others [1].

Currently, email has evolved into an archetypal medium [2] as they are branched into various branches like Yahoo mail, Google mail [4], Hot mail, Outlook [5], and so on. The services provided by them are entirely free of cost, for all web users who are adhering to some administration. Presently, email is considered as a secure international communication channel for its numerous functionalities. But once in a while, they can be very deceitful and hazardous because of a few "spam emails".

Normally, unsolicited emails or messages sent by individuals or organizations to consumers with the intention of illegal profiteering or sabotage are called as spam emails. The user's details are deceitfully collected on a webpage and later messages are sent to the domain's username. It's been twisted for monetary profits using the assortment of procedures and instruments that incorporate hoaxing, bonnets, open disinterested parties, mail transfers, mail instruments called mailers, then forth [8]. Filtering such harmful messages from legitimate ones may be a challenging undertaking for an assortment of reasons. Due to spam email, users or consumers experience several problems like exploitation of internet traffic, inadequate space for storing legitimate messages, condensed computational power, become a blockade for locating the additional email, waste their valuable time, and can also be a major security risk [9,10]. So, to make the electronic messaging framework safer and more effective, an apt system for filtering spam is very vital.

Spam filtering may be a method to find uninvited massage and stop moving into the user's inbox. Nowadays, varied systems are existing to engender a strategy for preventing uninvited bulk email. Most of the strategies have some inconsistency between false negatives and false positives that act as a barricade for many of the systems to create a well-off spam filtering system. Therefore a smart and operative spam-filtering system is the key mandate for all internet users.

II. Existing Methodologies

A. *The standard spam filtering method*

A significant number of spam filtering methods are existing today. Amid them, the method which makes use of a set of defined rules to identify whether an email is ham (not spam) or spam is the standard spam filtering method. A standard spam-based filter consists of the following steps [15-16]. Initially, several machine learning techniques are applied to check whether an email is a spam known as content-filters. Subsequently, data or information is mined from the email headers using header filters. Then the spam message is stopped if it is found to be existing in the blacklist file. Ensuing, some filter based on rules distinguish the sender through the subject line by using user-defined criteria. Following that, a permission filter is added to get the pre-authorization of the recipient before sending the message. Lastly, the filter using the challenge-response method is applied by getting permission from the sender to send the mail by applying an algorithm.

B. *Case-Based Spam Filtering*

Amongst various techniques, the most prominent technique for identifying spam emails is machine learning.

Comprehending that by decreasing the number of false-positive cases and increasing the quantity or amount of the training data, the accuracy improved significantly for classifying the Malay language spam Sharma et al. [14] described the Adaptive Approach for Spam Recognition. The article considered various Machine learning methods like Bayesian Networks, Random Forest, etc. and applied them on a single standard SPAMBASE dataset. The labeled spam/ham emails from a single account were grouped into a single entity and the accuracy was measured. The quality of this approach can be estimated by its accuracy, which was 85.65%.

C. Consumer-side and enterprise-level spam filtering

In this method, only by clicking through an ISP, the consumer will be able to send/receive emails [18]. There are some frameworks provided to filter spam at the consumer-level. An individual can install many such frameworks on his/her laptop for filtering spam. The consumer's or client's inbox is filtered by the framework by composing, managing, and receiving the messages which move along the mail user agent (MUA). To categorize spam emails in the network, a spam filtering framework must be installed on the mail server capable of interacting with the mail transfer agent. This system would provide additional efficiency by filtering the spam emails over the user's network; Principle-based grading is used by these frameworks. A principal-score is calculated for a message by using the pointers attached to the message. If the score exceeds the threshold value, then the email is considered to be spam.

III. NEURAL NETWORK APPROACH

An artificial neural network (ANN) is substantially used to solve a huge quantity of troubles in endless domain names of anthropological movement. Neural network fashions are utilized in fixing troubles which includes diagnostics, control, estimating, sample recognition, etc because of its standard approximation ability. Nevertheless, ANN's were most eminent in fixing classification issues. The following diagram illustrates how neural networks may be used to filter spam emails [3,10].

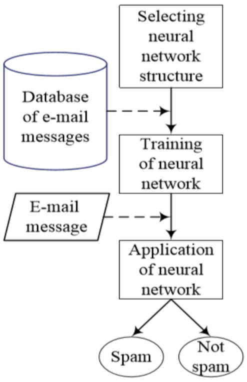
Selecting neural network structure
Database of e-mail messages
Training of neural network
E-mail message
Application of neural network
Spam
Not spam

Fig 3.1: Neural Network Technology for spam filtering

As depicted above, the usage of ANN technology for email messages classification entails the following steps:

• Choosing the shape of an ANN (i.e. choosing the appropriate architecture of the neural network like the number of hidden layers);

• Training the ANN model using training dataset ("spam"/"ham" emails);

The ANN can correctly identify spam messages and apprehend innovative sorts of spam in emails. The ANN's weights after the training phase contain information about the unsolicited emails that determine the performance of this type of application.

The following steps are involved in an ANN's filtration model [4]:

• finding email message facts, which include instances of "spam" and "ham" messages or emails;

• Initial information pre-processing and training pattern;

• ANN structure development: outputs, inputs, variety of layers and neurons in each layer;

• ANN training for spam email filtration version construction;

• ANN's unsolicited mail filtering model, followed by testing and evaluation of accuracy.

If the results of the evaluation model are not satisfactory, we need to return to the first step and carry out all the steps in the given order. The choice of the preliminary phase is decided experimentally.

IV. IMPLEMENTATION

Let us consider that the spam filter should be implemented with the help of Neural Networks for accurate classification of spam emails and this is a huge part of the application in the backend, and this application is presented to the users with the help of a suitable frontend in the form of a website for easier access of the users. The following steps are involved in implementation:

A. Data Collection:

Suitable large datasets containing most recent spam and ham emails should be collected from various sources. The number of spam emails collected should be of greater number for more precise calculations.

B. Pre-processing collected data:

The collected data should be processed into the required formats for easier analysis.

C. Feature Selection:

Tf-idf represents term recurrence opposite archive recurrence, and the tf-idf weight is a weight regularly utilized in data recovery and text mining. This weight is a factual measure used to assess how significant a word is to an archive in an assortment or corpus. The significance builds relatively to the occasions a word shows up in the report yet is counterbalanced by the recurrence of the word in the corpus. Varieties of the tf-idf weighting plan are regularly utilized via web indexes as a focal apparatus in counting and placing a report's significance given a client inquiry.

One of the only ranking capabilities is computed utilizing summing the tf-idf for each question term; many more sophisticated ranking features are versions of this easy model. Tf-idf can be correctly used for stop-words filtering in various situation fields including textual content summarization and classification.

Typically, the tf-idf weight is composed employing terms: the primary computes the normalized Term Frequency (TF), aka. The quantity of times a term occurs in a document, divided by way of the overall number of sentences in that document; the second one time period is the Inverse Document Frequency (IDF), computed because the logarithm of the range of the files within the corpus divided with the aid of the variety of files wherein the time-period seems unique.

- **TF: Term Frequency**: which measures how often a term occurs in a report. Since every document is exceptional in length, it's miles feasible that a term would appear lots more instances in long documents than shorter ones. Thus, the term frequency is often divided via the file length (aka. the entire variety of terms in the report) as a way of normalization:

  TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).

- **IDF: Inverse Document Frequency**, measures how critical a term is. While computing TF, all terms are taken into consideration are equally crucial. However, it is regarded that sure phrases, such as "is", "of", and "that", can also appear several times, however, they have little importance. Thus we need to overwhelm the recurrent phrases at the same time as scale up the infrequent ones, by:

  IDF(t) = log_e(Total measure of files / Number of records with t in it).

D. *Training the model:*

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

- The pre-processed email data is divided into the training and testing phase.
- Large data set is used to train the model to achieve more accuracy in predictions.
- The training data is further divided into a small amount for validation.
- The features of the data are collected using the tf-idf model.
- Now, our neural network is trained with the extracted features.
- Compare the accuracies of various models and determine that the neural network model is best suited.
- Test the model with the unseen test data.

E. *Prediction Phase:*

After the model is trained as specified in the previous phase, the model is ready for practical deployment. It accepts the users' specified features to predict spam emails in the future.

In the prediction phase, the tf-idf model is applied to the collected data and the features are generated. Then the test set of data is given to the model and the results are predicted.

The following flow chart is a step-by-step depiction of how the prediction phase would work for ANN's model.

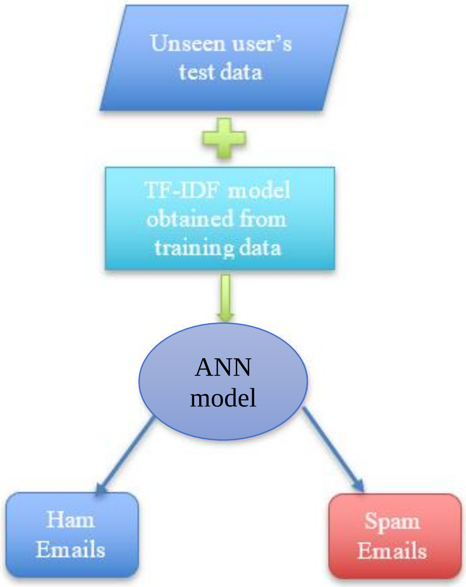

Fig 4.1: Flow chart depicting the prediction phase

F. *Accuracy comparison with popular ML algorithms*

For experimental purposes, we considered a standard spam email dataset known as the Enron dataset, which approximately consists of 50,000 emails. Of those emails, 35,000 emails are spam whereas the rest are ham. 40,000 random emails were used for training whereas the remaining were used for testing the model. We classify the same data using three other techniques as well: SVM, Random Forest, and xgboost. The calculated accuracies for all the techniques are tabulated in the following table

SL NO.	*ML Technique*	*Accuracy*
1.	SVM	93.45
2.	Random Forest	88.76
3.	Neural Network	99.76
4.	XGboost	88.50

Table 4.1: Experimental results

ACKNOWLEDGMENT

The conclusion of this paper brings with it a feeling of fulfillment, but it is never accomplished without thanking the individuals who're all answerable for its successful completion. Above all, we wish to convey our deepest sincere feelings of gratitude to my Institution, Sai Vidya Institute of Technology, for supplying me an opportunity to write this article.

We expand our unfathomable feelings of earnest gratitude to Dr. H S Ramesh Babu, Principal, Sai Vidya Institute of Technology, Bengaluru, for having us approved to perform the paper presentation on "SPAM EMAIL FILTERING USING ANN" successfully. We express our thanks to Prof. A M Padma Reddy, Director (A), Professor and Dean (Student Affairs), Department of Computer Science and Engineering, Sai Vidya Institute of Technology, for his regular enthusiasm. We express our earnest and genuine gratitude to K. Ananthapadmanabha, Professor and HOD, Department of Computer Science and Engineering, Sai Vidya Institute of Technology, Bengaluru, for his valued pointers and assistance. Finally, we would love to thank all of the Teaching, Technical school, and supporting personnel contributors of the Department of Computer Science and Engineering, Sai Vidya Institute of Technology, Bengaluru, for their guide.

REFERENCES

[1] A. Nazemi and M. Dehghan, "A neural network method for solving support vector classification problems," Neurocomputing, 152, pp. 369-376, 2015.
[2] A. Jantan, W.A.H.M. Ghanem, and S.A.A. Ghaleb, "Using modified bat algorithm to train neural networks for spam detection," Journal of Theoretical and Applied Information Technology, 95(24), pp. 6788-6799, 2017.
[3] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2, pp. 1097-1105, 2012.
[4] A.H. Mohammad and R.A. Zitar, "Application of genetic optimized artificial immune system and neural networks in spam detection," Applied Soft Computing Journal, 11(4), pp. 3827-3845, 2011.
[5] A.S. Katasev and D.V. Kataseva, "Expert diagnostic system of water pipes gusts in reservoir pressure maintenance processes," 2nd Int. Conf. on Industrial Engineering, Applications and Manufacturing, 7911651, 2016.
[6] A.S. Katasev and D.V. Kataseva, "Neural network diagnosis of anomalous network activity in telecommunication systems," Dynamics of Systems, Mechanisms and Machines, Dynamics, 7819020, 2016.
[7] A.S. Katasev, D.V. Kataseva, and L.Yu. Emaletdinova, "Neuro-fuzzy model of complex objects approximation with discrete output," 2nd Int. Conf. on Industrial Engineering, Applications and Manufacturing, 7911653, 2016.
[8] C. Wilcox, C. Papadopoulos, and J. Heidemann, "Correlating spam activity with IP address characteristics," Proceedings - IEEE INFOCOM 5466660, 2010.
[9] C. Ge, B. Wang, X. Wei, and Y. Liu, "Exponential synchronization of a class of neural networks with sampled-data control," Applied Mathematics and Computation, 315, pp. 150-161, 2017.
[10] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 6248110, pp. 3642-3649, 2012.
[11] E. Volna, T. Sochor, C. Meli, and Z.K. Oplatkova, "Soft computingbased information security (Book Chapter)," Multidisciplinary Perspectives in Cryptology and Information Security, pp. 29-60, 2014.
[12] G. Goswami, R. Singh, and M. Vatsa, "Automated spam detection in short text messages," Advances in Intelligent Systems and Computing, 390, pp. 85-98, 2016.
[13] I.V. Anikin and R.M. Gazimov, "Privacy preserving DBSCAN clustering algorithm for vertically partitioned data in distributed systems," Int. Siberian Conf. on Control and Communications, 7998473, 2017.
[14] J.L. Castro, C.J. Mantas, and J.M. Benítez, "Neural networks with a continuous squashing function in the output are universal approximators," Neural Networks, 13(6), pp. 561-563, 2000.
[15] K. Esfandiari, A.A. Ghoreyshi, and M. Jahanshahi, "Using Artificial Neural Network and Ideal Adsorbed Solution Theory for Predicting the CO2/CH4Selectivities of Metal-Organic Frameworks: A Comparative Study," Industrial and Engineering Chemistry Research, 56(49), pp. 14610-14622, 2017.
[16] L.Y. Emaletdinova and E.D. Tsaregorodtseva, "Algorithms of constructing a neural network model for a dynamic object of control and adjustment of PID controller parameters," Russian Aeronautics, 56(3), pp. 247-256, 2013.
[17] L.Y. Emaletdinova, I.V. Matveev, and A.N. Kabirova, "Method of constructing a neural regulator for the automatic one-dimensional control of a technical object," Russian Aeronautics, 58(2), pp. 227-232, 2015.
[18] L.Y. Emaletdinova and A.N. Kabirova, "Neural fuzzy controller to control the angle of heel and the course of the unmanned aerial vehicle," Dynamics of Systems, Mechanisms and Machines, Dynamics, 7819001, 2016.