

# A New Approach to Email Auto Responder

Athul Krishna P.B <sup>[1]</sup>, Raheenamol M.R <sup>[2]</sup>, Jesna M.S <sup>[3]</sup> and Ms. Sheena Kurian K <sup>[4]</sup>

BTech Student <sup>[1],[2] & [3]</sup>, Associate Professor <sup>[4]</sup>

Department of Computer Science & Engineering

KMEA Engineering College, Ernakulam

Kerala - India

## ABSTRACT

Email plays a vital role in all organization whether it is business, government, manufacturing or education considering the use of email as a primary communication tool in companies for manufacturing. Large companies assign staff for replying mails to the customer's queries. It creates a large substantial problem for an employee to read and reply to the email sent by customers. As a solution to this problem, our paper suggests a method through which we can train the machine automatically to reply to the customer.

**Keywords** :— Artificial Neural Network, Classifiers, Lemmatization, Lowercasing, Stemming, Tokenization.

## I. INTRODUCTION

In highly reputed computer manufacturing companies and other organizations the customer interacts with the manufacturers through email as the standard means of customer services. Customers are the great assets for a company. So, in every company, there is a department which handles customer services. They provide services or solutions to the customer's problem. Customers send their product issues to the company through email, and the customer representatives read and reply to the email. Efficient reply is provided by the representatives to the customers; This method is effective if there are large number of employees to attend the customers or the number of customers should be less. Large companies assign staff for reading and replying mail to the customers. If more emails are received, then more employees are needed to process the mail. More money and time are consumed. A case study reported that a firm in the UK with 2850 email users actually spent £40,848 per day for reading and responding to emails [1]. Solution to this process is provided by AERS.

In AERS the customer interacts through their login, for example, a computer manufacturer gets a complaint on the installation of software. When the customer faces such a problem, they send a mail to the company. The solution for the complaint will be present in the dataset which is redirected to the customer automatically. So in order to provide an efficient and effective way to process emails we propose automatic email responding system.

## II. LITERATURE REVIEW

Weiwen Yang and Linchi K Wok [2] proposed a method for improving the automatic email responding system for computer manufacturers via machine learning. A customer may have different queries, these queries are passed to the system. In his paper he classifies the email into multiple categories by machine learning algorithms and is stored in a database, where the database contains the web pages link to

each category and predefined solution to each answer for each category. They use K-means++, KNN and NB algorithm for categorization of email into multiple categories. The email received from the customers are passed to classifiers and partitioned into predefined categories. Once the category of email is identified, the classifier sends the emails with the category to the selector. When the selector receives the notification from the classifiers, they pick the correct solutions from the database and redirect to customer's email from the company server.

Judy Strauss et al [3] proposed an idea about the reaction of management to the customer. His study deals with exploring company's responses to genuine complaints via email and customer reactions to these responses. The method behind his ideology is to collect the complaints from customers and send this email letter to the researcher. A content analysis was performed on company responses to complaint email using predetermined classification and characteristics. The response rate is measured and time stamp is calculated in order to reply the email.

## III. PROPOSED METHOD

This paper suggests a method of automatic email responding system using a neural network. The customer sends mail to the company about the product. This complaint will be collected in the company database and the matching answer will be sent to the customer via the company's email id. We use the concept of the neural network for training and prediction of email. Fig 1 shows customer and system interaction; the customer sends queries and the system automatically replies to it. Fig 2 shows the detailed architecture of the system. The system consists of preprocessing, vectorization, model creation & training, evaluation and solution model. The preprocessing phase consists of word tokenization, lowercasing, stemming and lemmatizing. The tokenized word is then vectorized to form bag of words where the stopwords are removed. The result

obtained will be used for model creation and is trained using neural network.

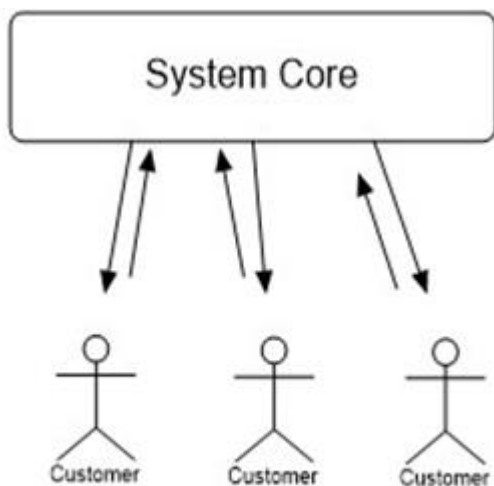


Fig 1: Basic Architecture

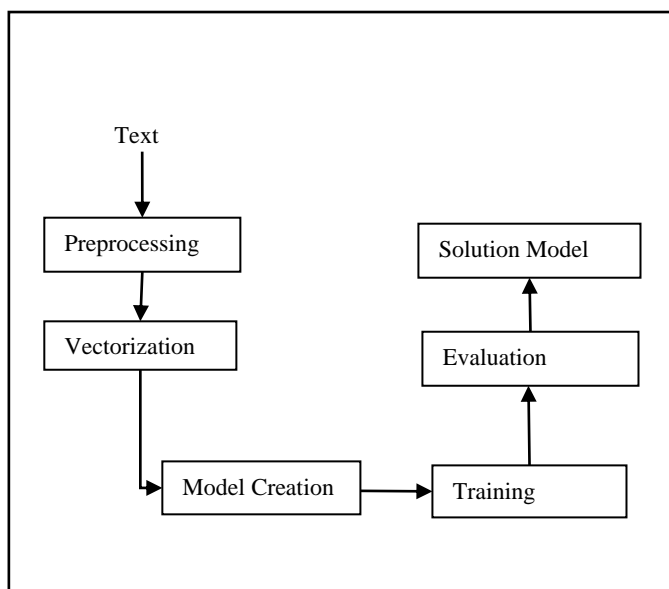


Fig 2. Detailed Architecture

#### IV. PREPROCESSING

Preprocessing refers to the transformations applied to our data before feeding it to the algorithm. Data preprocessing is a technique that is used to convert the raw data into clean data set. The real data will be inconsistent, incomplete and lacking in certain behavior's or trends and is likely to contain many errors. So preprocessing techniques is used to resolve such issues. The need of preprocessing is that to get a better result from the applied model [8].

The preprocessing techniques applied to the text are Lowercasing, Tokenization, Stemming, Lemmatization.

##### A. Word Tokenizing

The Tokenizing is a process of splitting tokens from sentences or paragraphs. Tokens include keywords, words, and phrase. During tokenization, punctuations are discarded. This token becomes the input for the next process. The ending point of a word and the beginning of the next word is called word boundaries. Tokenization is also known as word segmentation. [4]

Example of Tokenization: Sentence: -“This is an apple” will be broken into tokens “This”, “is”, “an” and “apple” after tokenization.

##### B. Lowercasing

In lowercasing, each single token of the input text is lowercased. It is the simplest preprocessing technique and has become a popular practice in deep learning library modules and word embedding packages. [9]

Example of Lowercasing:

“Oreo supports new emoji that where included in the Unicode 10 standard.” In this case, lowercasing may have negative impact due to ambiguity. Since Orio Android in our example and Oreo biscuit would be considered as identical entities.

##### C. Stemming

Stemming usually refers to a process that removes ends of words to achieve the goal correctly. In stemming it produces morphological variants of a base root. In stemming, errors occur when two words are stemmed to same root word that is of different root or when two words are stemmed to different root word that belongs to the same root word. [5]

For example, the word ‘goes’, ‘going’ and ‘gone’ will be stemmed under the root ‘go’. Stemming algorithm has the following steps:

1. Import PorterStemmer from nltk.stem.
2. Spilt the input text into tokens.
3. Pass each token through PorterStemmer.stem() method.

##### D. Lemmatization

Lemmatization is the process of grouping together the different inflected form of the word and analyzing it as a single item. Stemming and lemmatizations are not similar, but we prefer lemmatization over stemming because morphological analysis of the word is done by lemmatization and it also brings context to the word. Both stemming and lemmatization can be done using Natural Language Toolkit (NLTK). [5] Lemmatization algorithm has the following steps:

1. Import WordNetLemmatizer from nltk.stem.wordnet.
2. Split the input text into tokens.
3. Pass each token through the method WordNetLemmatizer.lemmatize().

## V. VECTORIZATION

In vectorization we can represent words in the vector form. We use bag of words technique for vectorization. Firstly, we create bag of words for the dataset and save the words obtained in the technique as a pickle file. The words present in the bag of word will be represented as 1 and 0 for absent. It actually maps the textual data to real valued vectors.

### A. Bag of Words (BOW)

It is the basic model used in natural language processing and information retrieval. The order of words in a document is discarded and tells whether the word is present in document or not. BOW is used to extract the features of the document. It counts the occurrence of each words in a document and take the word count as the feature.

Example of Bag of words: Sentence: - “John likes to eat icecream. Akash likes ice cream too.”Based on the sentences a list is constructed or sentence is tokenized as:

“John”, “likes”, “to”, “eat”, “icecream”, “Akash”, “likes”, “icecream”, “too”

Bow=

{“John”:1,“likes”:2,“to”:3,“icecream:2”,“Akash:1”,“too:1”}

## VI. MODEL CREATION AND TRAINING

A model is created which can be used to predict the probability of the word in the sequence, based on the observed sequence. The model predicts the probability of each word from the given input.

A classifier act as a model to predict the class of a given data. We train the classifier using a certain amount of data and future prediction is done on the basis of the trained data. There are different types of classifiers in machine learning for classification. In the Weiwan Yang [1] paper he used Naïve Bayes (NB) model, K-Means++ and K-NN.

We propose to train the model using Artificial Neural Network (ANN) shown in Fig 3. ANN works similar to our biological neuron i.e., a simple neuron consists of the cell body and a number of protrusions called dendrites, and a single branch called the axion and communicate each other through a point called synapse. [6]

An artificial neural network is a weighted directed graph consist of a set of input and output. Each input connection has a weight attached to it. Each node in the network is called a neuron. During the training phase, we train the network by adjusting weight so that the network will be able to predict the correct class label for inputs.

The Neural Network is one of the learning algorithms used in machine learning. The network consists of a hidden layer between the input and output layer. Hidden nodes are the nodes in the hidden layer that comes between the input layer and the output layer. The activation function is applied to the hidden layer that consists of hidden nodes. The positive or negative neurons either inhibit or excite the input. The activation function is used to control the transportation of neurons [7].

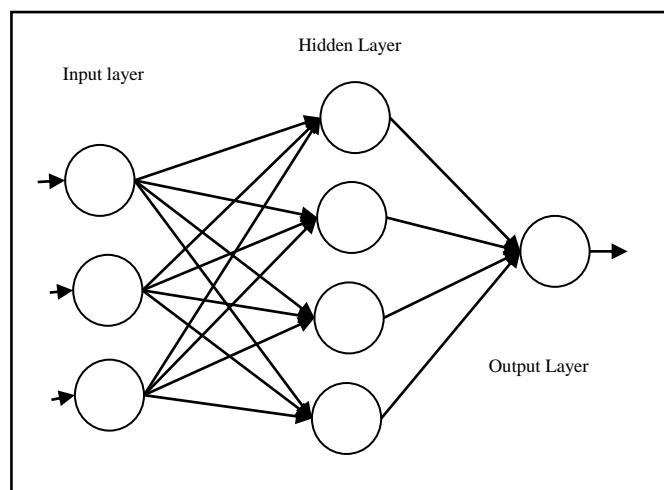


Fig 3. ANN Architecture

In training phase, the dataset is classified to get the output. The output is determined with predefined data. This process is done by using the backpropagation algorithm.

Steps involved in the artificial neural network:

1. Assign random weights to all linkages to start the algorithm.
2. Using the inputs and the linkages (hidden nodes) find the activation rate of hidden nodes.
3. Using the activation rate of hidden nodes and linkages to output, find the activation rate of input nodes.
4. Find the error rate at the output node and recalibrate all the linkages between hidden nodes and output nodes.
5. Using the weights and error found at the output node, cascade down the error to hidden nodes.
6. Recalibrate the weights between the hidden node and the input nodes.
7. Repeat the process until the convergence criterion is met.
8. Using the final linkage weight score the activation rate of the output nodes.

The neural network can be implemented using tensorflow. TensorFlow is computational frame work for building machine learning model. It provides variety of toolkit that helps to construct the model at preferred level of abstraction. In training phase, the first step will be to fetch the labeled training data of positive and negative sentences and is processed with its vectors that is then used for training and testing datasets. Then placeholders are kept for input and output networks. These networks perform convolutions over the ordered word vectors. TensorFlow is used when the pool layer is added with convolution. The pool layer is used to extract most significant elements in each convolution. Finally, the fully connected layer is added in such a way that non-

linearity to the model is added. Then the resulting output is classified using certain functions.

#### IV. EXPECTED RESULT

The matching response is provided to the customer automatically. The classified query is fed to the evaluation stage where accuracy is checked and a predefined response is collected and is passed to the customer.

The expected result is shown in Fig 4, the response is provided to the customer automatically from the corresponding representatives.

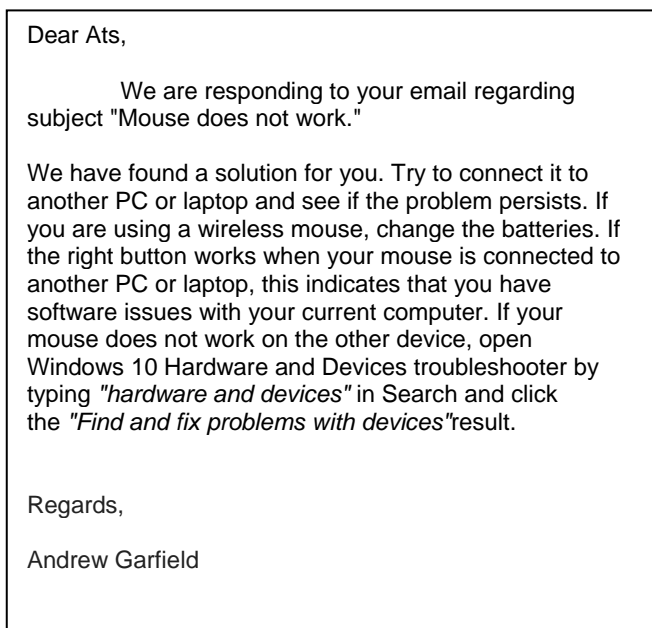


Fig 4: Sample Email Reply

#### V. CONCLUSIONS

We have presented an overview of the proposed solution to the automatic email reply using a neural network. This response system can classify the emails and send replies to the customer based on their mail. The information from each individual is analyzed by extracting the important words and classifies it using Artificial Neural Network and the resultant answer to the mail is identified from the database and is redirected to the customer.

Algorithms are used to classify the text in the email and provide the answers automatically. The algorithm performs a classification of emails based on the content with the help of information extraction. Email reply can be wrong due to unintended situations. So, step by step approach is done to provide an accurate reply. Many different types of emails are trained in order to provide an accurate answer. So, our aim is to provide a suitable and efficient email response system.

#### REFERENCES

[1] T.W. Jackson, A. Burgess, and Edwards, "A simple approach to improving email communication: Going

back to basics, "Communications of the ACM", 2006, June, vol. 49, no. 6, 107-109

[2] 2012 International Conference on Information Management, Innovation Management and Industrial Engineering "Improving the Automatic Email Responding System for Computer Manufacturers via Machine Learning" Weiwen yang school of engineering and applied science Columbia University New York, Usa wy49@columbia.edu

[3] Judy Strauss Donna J.Hill 2001 John Wiley & Sons, Inc. and direct Marketing Educational Foundation, Inc. "Consumer Complaints by Email : An Exploratory Investigation of Corporate Responses and Customer Reactions" Journal of Interactive Marketing Volume 15/ Number 1/ Winter 2001

[4] Fco. Mario Barcala, Jesus Vilares, Miguel A. Alonso, Jorge Grana, Manuel Vilares "Tokenization and Proper Noun Recognition for Information Retrieval" Departamento de computacion, Universidade da Coruna Campus de Elvina s/n, 15071 La coruna, Spain

[5] Mayuri Mhatre, Dakshata Phondekar, Pranali kadam, Anushka chawathe, Kranti Ghag "Dimensionality Reduction for Sentiment Analysis using Pre-processing Techniques" Information Technology Department SAKEC, chembur, Mumbai, India, IEEE 2017

[6] Anil K. Jain Michigan State University "Artificial Neural Networks: A Tutorial "

[7] "How does Artificial Neural Network algorithm work?" [Online] Available: "https://www.analyticsvidya.com/blog/2014/10/ann-work-simplified/ "

[8] "Data Preprocessing for Machine learning in python"[Online] Available: "https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/

[9] Jose Camacho-collados and Mohammad taher Pilehvar "On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis" School of Computer Science and Informatics Cardiff University and Iran University of Science and Technology.