Chapter 1

Introduction

1.1 Introduction

We live in the era of modern technology. Here almost every things are dependent on the technology .People are getting use to the technology to make their life easy and more comfortable. Today's technology is mostly an advancement of old technology, the influence of technology in modern life is immeasurable, we are using technology in various ways and sometimes the way we execute different technologies ends up offending our lives or the society. What we say modern technology is technically not that new in most cases. For instance, communication technology has developed with years, right now we use email which has been an improvement of Fax.

Email is broadly used as a way of business communication and it is a highly effective communication tool. Email is mostly inexpensive, it only requires an internet connection that is already present in the business. In the result as online social networking and communication is appealing to the public. Statistics indicates an improvement of 3% in the number of Email users with an average of 1.7 from 2011 to 2015. Email accounts per user counted in 2015. Moreover, business Email communication accounts for the majority of the total Email with over 108.7 billion Emails exchange daily, and Email remains the most common process of business pace communication.

In order to exchange the mail most of the large private company use private mail server. They provide all their employees an individual email account. And continue the communication with them. Using a private mail server, the biggest problem is to handle the spam challenge. There are some tools that can handle the problem also.

But there exist another problem that if any of the company employee is doing the conspiracy about the company, exchanging any sensitive information that can make a bad effect for the company, no way to detect it. There exist some big named company once that spiraled downward into bankruptcy due to the conspiracy between their employees. And this problem is getting increased day by day. Now mail is the most efficient way of transferring the information between the people.

In this study, we propose a system that will automatically detect the conspiracy related mail from real time mail box. As the detection of conspiracy will be fully automated, account of the sender and receiver employee will be detected and the information will be safe. We have to face some challenge in this thesis. Collecting real-time mail from mail server by customizing the POP3 protocol and at the same time analyzing them to detect conspiracy will be challenged. Also we have to first build an algorithm which will be able to detect conspiracy from textual content. It is the biggest challenge for us.

1.2 Previous Works

The most common process to text sentiment analysis consist of detecting the frequency of features (words) of given positive or negative semantic value. In that way, sentiment analysis has a lot in similarity with classical text classification and mining [1], and any one would be enticed to implement statistical keyword significance metrics, for example chi-square or TFIDF. Unfortunately, these do not provides expected result for sentiment classification. For better result some work has been done using standard machine learning techniques to sentiment analysis, such as Pang and Lee [2] who used Bayesian classifiers, maximum entropy and SVM. Similarly, Turney and Littman [3] used latent semantic analysis (LSA) to calculate the relationship between words observed in a text and a predefined praise word. But the unique character of the challenge of sentiment classification has given improve to innovative new technique as well.

There are some works on analyzing email data. Some of these tried to analysis the large data of email. They use sentimental analysis to detect positive negative sentiment. Sisi Liu and Ickjai Lee has proposed a framework for sentiment analysis in Email using a hybrid scheme of process combined with Kmeans clustering and a classifier , support vector machine. The measurement for the design is conducted through the correlation among three labeling methods, SentiWordNet labeling, Kmeans labeling, and Polarity labeling, and there were five classifiers, including Logistic Regression, Decision Tree, OneR and Support Vector Machine, Naïve Bayes, [4].Combination of clustering approach with SWN lexicon for sentiment analysis in blogs was proposed by Feng, Wang, Yu, Yang and Yang [5]; Li and Wu [6] proposed Kmeans clustering for hotspot indication

and SVM sentiment classification and prediction. Current analyzation on text mining and sentiment analysis mostly addresses large scale social media data, such as Twitter corpus, blogs and Facebook data. For instance, Li and Wu [6] proposed a study on hotspot finding through online forum.

Moreover, Balasubramanyan, Routledge and Smith [7] propose an algorithm for the prediction of poll results by using public opinion. As for Email data classification, most studies focus on the identification of spam emails, discarded emails, the study of social networking among mails, and priority issues. [8] [9] [10]. However, few research has been conducted on mail conspiracy analysis. Mohammad and Yang [11] analyzes the gender classification of sentiment axis among set of sentiment labeled mail data, and Hangal, Lam and Heer [12] proposed a system for visualizing archived mail data with sentiment words. Other than these studies, a systematic and structured design for conspiracy detection from Email data has not been investigated yet.

1.3 Present State and Contribution

The goal of this project is to design a system to detect conspiracy from the Email conversation between the employees of any company or farm. This system will detect the suspicious conversation between the employees against the company in deferent angle. We will detect the sentiment between the test conversations. Here we are proposing the method of designing a system that can automatically analysis the conversation and give the feedback with the related employee name to the owner or the authority

1.4 Motivation

Enron was an energy company in American based on Houston, Texas created by Ken Lay. This company became bankrupted in October 2001. It was the largest bankruptcy organization in the history of American at that time. Enron was accused as the biggest audit failure.

Many executives at Enron were blamed for a variety of charges and many were later sentenced to prison. Many of them found guilty of unauthorized destruction of documents relevant to the SEC

investigation, that voided its license to audit general companies and effectively closed the farm. During the investigation Federal Energy Regulatory Commission made the email data of these employee public. After reading about the Enron case study something got to mind to make something that can automatically investigate the email data that are passing through the employees. Thus I was interested in analyzing data by classifying them into conspiracy class.

1.5 Prospects

This project has a large prospective in the present word. It has a practical value in any organization where individual exchanges confidential information among themselves. It will audit the information exchange. It will help the management to have a good look over the employees for not being harmed. It will make sure the proper working condition in the work place. It maintain the commitment and trust between the individual and confirm the profit of the company. It provides the company to relay on their strategy and model of work properly. This study will help a company to monitor and predict any upcoming fall or loss in many ways. Any working place can analyze their entire system by implementing this project in their system. It also proposed a model and framework to convert a psychological theory into a machine. It is also an important aspect of research. So we can say that this thesis and proposed model of system can make a good impact on the digital automated world.

1.6 Objectives

The study will be carried out to achieve the following goals specified:

1. To develop a framework for extracting the content of email from the email data.
2. To develop a framework for conspiracy detection by analyzing the extracted email contents.
3. To send the verdict message to the owner of the company regarding conspiracy
4. To implement a mail server with public IP for real time email communication.

1.7 Organization of the Project

This report is arranged with six chapters. Chapter one contains some introductory text and preliminary information about our work, previous works contains the similar forms of work that has been worked before, present state and contribution contains my contribution in this work ,motivation of the research specify the initial thought that makes me interested in this work . Chapter two contains literature review about the required knowledge about the project and gives the over view of the technique and study that should be done by me.

Chapter three deals with the overall process of the system and my working method or suggested technique and procedure. In chapter four, we have presented our implemented work and in chapter five, experimental results and evaluations are explained. Finally, chapter six concludes our overall work.