Chapter 1

Introduction

1.1 Introduction

We live in the age 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. Modern technology is simply an advancement of old technology, the impact of technology in modern life is unmeasurable, we use technology in different ways and sometimes the way we implement various technologies ends up harming our lives or the society we leave in. What we call modern technology is technically not so new in most cases. For example, communication technology has evolved with years, nowadays we use email which has been an advancement of Fax.

Email is widely used as a form of business communication and overall it is a highly effective communication tool. Email is inexpensive, only requiring an internet connection that is generally already present in the business. As a result as online social networking and communication is increasingly appealing to the public. From 2011 to 2015, statistics indicates an increase of 3% in the number of global Email users with an average of 1.7 Email accounts per user counted in 2015.Furthermore, business Email communication accounts for the majority of the total Email traffic with over 108.7 billion Emails exchange every day, and Email remains the most common way of business workspace 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 approach to text sentiment analysis consists in detecting the occurrence of features (words) of known positive or negative semantic value. In that sense, sentiment analysis has a lot in common with classical text mining and classification [1], and one would be tempted to use statistical keyword significance metrics such as chi-square or TFIDF. Unfortunately, these do not give good results for sentiment classification. To be sure, some work has been done applying standard machine learning techniques to sentiment analysis, such as Pang and Lee [2] who used Bayesian classifiers, maximum entropy and SVM. Likewise, Turney and Littman [3] used latent semantic analysis (LSA) to measure the relationship between words observed in a text and a predefined praise word set. But the unique nature of the challenge of sentiment mining has given rise to innovative new approaches 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 Email sentiment analysis using a hybrid scheme of algorithms combined with Kmeans clustering and support vector machine classifier. The evaluation for the framework is conducted through the comparison among three labeling methods, including

SentiWordNet labeling, Kmeans labeling, and Polarity labeling, and five classifiers, including Support Vector Machine, Naïve Bayes, Logistic Regression, Decision Tree and OneR[4]

Feng, Wang, Yu, Yang and Yang [5] combine clustering approach with SWN lexicon for blogs sentiment analysis; Li and Wu [6] utilize Kmeans clustering for hotspot detection and SVM sentiment classification and prediction. Current research on text mining and sentiment analysis mainly addresses large scale social media data, such as Facebook data, Twitter corpus, and blogs. For instance, Li and Wu [6] conduct a study on hotspot detection through online forum.

Additionally, Balasubramanyan, Routledge and Smith [7] propose an algorithm model for the prediction of poll results using public opinion mining. As for Email data analysis, most studies focus on the identification of spam mails, discarded mails, the study of social networking among Emails, and priority issues. [8] [9] [10]. However, less research has been conducted on Email conspiracy analysis. Mohammad and Yang [11] study the gender difference in sentiment axis among set of sentiment labeled Email data, and Hangal, Lam and Heer [12] design a system for visualizing archived Email data with sentiment words tracking. Despite of these studies, a systematic and structured framework 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 American energy company based on Houston, Texas created by Ken Lay. This company became bankrupted in October 2001. It was the largest bankruptcy reorganization in American history at that time. Enron was cited as the biggest audit failure.

Many executives at Enron were indicted for a variety of charges and some were later sentenced to prison. Many of them found guilty of illegally destroying documents relevant to the SEC investigation, which voided its license to audit public 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. So we can say that this thesis and proposed model of system can make a good impact on the digital automated world.

1.6 Organization of the Project

This report is organized into 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.