

Chapter 5

Experimental Results

We have tested our system with extensive experiment. In this section, we first introduce how data are collected for our Conspiracy Detection Model. Then we will present the performance of the system and compare it with existing systems.

5.1 Data Collection

For Conspiracy detector we collect the data from the real environment. First of all finding the email data is not so easy. There is only some of real email dataset available in this world that are free for general research. We use Enron Dataset. So many researches have been done successfully with this set of data.

5.1.1 Green data collection

As we are going to use data of two classes. So far our plan was to collect the official email like office affair, work related email, deal related, client related, gratitude related, personal email, internal component operation, legal advice, humor, friendship affection related, jokes, forwarding email, Logistic arrangement etc. We have used the data with perfectly classified with these sort of classes. These sort of data or email are frequently exchanged between the employees of the company. So we have to classify these data as a green data.

We collect the email from that dataset and labeled it with class 0(zero). These are our Green dataset. We have stored these data in a csv file with two rows. Row one is for the body of the email. And the second row is for the sentiment or class. Class row holds the zero as the sentiment.

1	internal projectJeff I have prepared the attached chart for you. It captures the ideas we discussed last week.Please call when you have a chance. I look forward to talking with you	0
2	KarenHere it is!Plenty of good Houston input here as well as Europe.An excellent general article has resulted; nice pictures too. I'll get some original copies couriered over to you as	0
3	Jim- Do you have a list of who I talked to in Houston and their affiliation. I would like to keep straight who I have spoken with. The only card I received is from Michael Geffroy. Let	0
4	Once the website is functional, it will be helpful to have a complete history of deals and having these kind of memos in dash format will aid this process. One final thought, Michael	0
5	To keep pace with the fluid and fast-changing demands of our equity trading activities, Enron Wholesale Services ("EWS") has recently revised its official Policies and Procedures Reg	0
6	If you have already certified compliance with the Policies and Procedures during the 2001 calendar year, you need not re-certify at this time, although you are still required to to re	0
7	You are required to become familiar with, and to comply with, the Policies and Procedures. The newly revised Policies and Procedures are available for your review on LegalOnline	0
8	You must certify your compliance with the Policies and Procedures within two weeks of your receipt of this message. The LegalOnline site will allow you to quickly and convenient	0
9	Attached is Enron North America Corp.'s suggested revisions to your pro forma confidentiality agreement. Please review and let us know if the suggested revisions are approved.TI	0
0	Outsource for SuccessEnjoy management flexibility and the benefits of a secure,carrier-class environment with Sprint E Solutions Web hostingand collocation services. Learn about	0
1	Many Linux vendors have released a patch for the xinetd package that fixes a flaw in the way the application deals with TCPWAIT commands. The problem prevents the linuxconf-w	0
2	Linux-Mandrake has issued a patch for its tcpdump package thatfixes a potential buffer overflow vulnerability. The flaw couldbe used in a remote attack on the tcpdump process. T	0
3	According to an alert from Linux-Mandrake, several flaws havebeen found in the UW-IMAP package that could allow anauthenticated user to gain greater shell command access. T	0
4	To subscribe or unsubscribe to any Network World e-mailnewsletters, go to:http://www.nwnewssubscribe.com/news/scripts/notprinteditnews.aspTo unsubscribe from promotional e	0
5	FYI, the following website describes a variety of econometrics methods - which we plan to implement to solve the private firm problem. Many of these methods were briefly desc	0
6	Martin,You may find it useful.Vince	0
7	This is not the most clearly drafted provision as it was the subject of much negotiation....and was significantly narrowed from the original scope proposed by AEP. Thus, please con	0
8	This appears to be supply deal for deliveries at Katy that would be subject to the 90 day non-compete. Sandi is in the process of getting me a copy of the non-compete so that I can	0
9	Dan,Please find below details of Phy Gas deals with GTC agreements:	0
0	Please let me know if you need any more details.Thanks,Richardx54886	0
1	You are required to become familiar with, and to comply with, the Policies and Procedures. The newly revised Policies and Procedures are available for your review on LegalOnline	0
2	Enron Wholesale Services ("EWS") maintains official Policies and Procedures Regarding Confidential Information and Securities Trading ("Policies and Procedures"), which have been	0
3	You must certify your compliance with the Policies and Procedures within two weeks of your receipt of this message. The LegalOnline site will allow you to quickly and convenient	0
4	Whatever the explanation, the plain fact is that FERC and theadministration have yet to offer California any significantrelief.	0
5	Martin, LanceWhat do you think?Vince	0
6	As we discussed during our dinner, I think the two biggest sources> of benefits from re-structuring will come from getting the demand-side> involved in the market and from more	0

Fig 5.1: Green Data CSV file.

5.1.2 Red data collection

In this section we are going to explain the process of collecting the red data. It was not as easy as collecting the green data. As we are designing a model that can easily detect or predict the conspiracy in the email data. We have to learn the machine about the sentiment and psychology behind the concept of conspiracy. We have to frame the related word and concept clear about the theory.

As conspiracy is a psychological concept in human life. First we had to study through the concept of the conspiracy theory and about the working place conspiracy theory. There are too many conspiracy theory over the world. But we have just look through the working place conspiracy theory and the study over the theory. We have collected the consequences of conspiracy theory and the reason behind it. After all that study we made a list of situation based on the theory. We have come to the concept that there could be 3 possible angle of conspiracy in a working place that may causes the after effect that we have mentioned earlier. So we have sorted some point in which we will give our focus to create the real time environment and find the email with the concept of conspiracy.

Here we came out with the three angles of conspiracy, they are:

- 1 **Financial conspiracy**
- 2 **Organizational conspiracy**
- 3 **Reputational conspiracy.**

5.1.2.1 Financial Conspiracy

It reflects the concept of harming a company financially planning with the employees of that company in several way. It could be with direct fraud in financial account, could be investing in any dead project, missing the proper paper work in every financial transaction in the office place

5.1.2.2 Organizational Conspiracy

This concept reflects the view of overpower the company from the current management or owner, chairperson. This means the organizational change as well as the leadership changes in between the company. It could be in various scale of changes.

5.1.2.3 Reputational Conspiracy

It means the overall reputation of the company such as any rumor about the management, pricing, strategy, working condition, and financial statement so many fake news can harm the reputation of the company. This types of conspiracy can effect a company over night and destroy the socio-economic value of that company

After reviewing these concept more and more we select some people who can visualize things in real life. We have given them the knowledge about the theory in every angle with some example of conspiracy related conversation. We have cleared the concept of that individual with every way they asked. After some grooming, these person was monitored by us during making the data as form of email. We monitor the data about the reflection of the concept through the email properly. And after a certain period of time we became successful to make some conspiracy reflecting emails.

That was a great success for us as we at last have something to teach the machine. Then we collectively do this process in various environment. And collect these data. We have used almost 30 different persons to collect these data. So that was the most challenging part in collecting data set.

As we are trying to teach the machine a purely psychological concept of human nature, it was a tough job for sorting out the reflection of conspiracy throughout the email. After collecting these data we have made a csv file leveled with one in the sentiment column, the file looks like the figure

5.2

Fig 5.2 will reflect the Red email in a csv file:

Replace committee.	1
As you know,everyone is suspecting the CEO of this company for this collapse.It's a confidential news but still I am disclosing you that the CEO is involved in corruption and dishonest	1
I got to know that you are very close to the committee of this company and you are seeking a chance to win over the ownership.I have a great offer for you.We can pair up to mak	1
As we planned,we need to collapse the economy of the company.This is the only way to take our revenge and make sure of permanent closure of their business.	1
Can we meet tomorrow? I need to discuss some secret affairs with you.Actually we can crash the system.This will weaken the whole management as they will be in a fix.These kind	1
We must work together as a group to take control of this bank. Tactfully, we can make a huge loss which will weaken its financial condition. Moreover, the market price of share cai	1
It is came to know by rumour that XYZ bank has gone to bankruptcy. You also may aware of this issue.The financial condition of this bank is weakening day by day. As a CEO of the r	1
I am going to sell company shares to my friends at a cheap rate to drastically reduce company share prices.	1
I have heard that MR. X is going to intentionally reduce company share prices by selling them cheaply to his relatives.	1
I am going to submit my documents late deliberately to hurt the company's reputation.	1
Mr. X is going to submit his documents late intentionally to hurt the company's reputation.	1
We are going to make MR. X our partner though he is unsuitable for the position.	1
I am going to omit some details in my next report.	1
I have heard that MR. X is going to omit some vital details in his next reports.	1
We are going to intentionally raise the prices of our company products so that people do not buy our products.	1
I have heard that MR. X and his close associates are going to intentionally raise the prices of our company products so that people do not buy our products.	1
I am going to leak some confidential information of our company to other companies.	1
I have heard that MR. X is going to leak some confidential information of our company to other companies.	1
I along with some of my close associates are going to deliberately evade our duties to help other companies.	1
MR. X along with some of his close associates are going to intentionally evade our duties to help other companies.	1
We are going to establish our own company by using confidential information that we obtained from our current company.	1
I heard that MR. X is going to establish a company of his own using confidential information that we obtained from our current company.	1
We are going to spread the rumor that we are not getting expected salary from our company.	1

Fig 5.2: Red Data set

5.2 Evolution of the System

The collected data are used to evaluate the system. The system is evaluated on the basis of two ways:

1. Evaluates from the mail dataset by splitting them into training and testing
2. Evaluate the system in real time.

5.2.1 Evaluates from the mail dataset

We have done this work in our algorithm during training the model. In that period we have separated the dataset with training set 80% and testing 20%. And after that we have a certain accuracy of ~68%

We have used 450 class 0 tagged data to evaluate the detecting percentage of the model. It gives us that 324 email with green detection and 126 email with false detection.

Thus the accuracy over detecting the green data is $\frac{324 \times 100}{450} = 72$

So the accuracy in the Green data detection is 72%

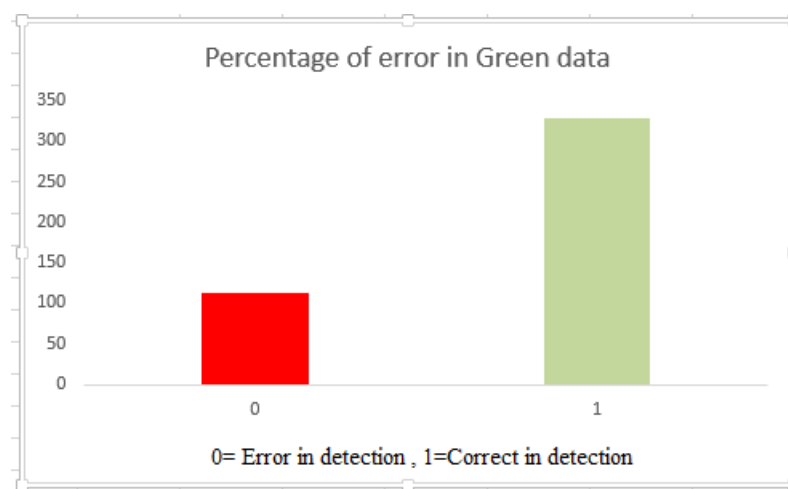


Fig 5.3: Percentage of error in Green Data

Again in the Red data we continued the process of detection. Here we also uses 450 email data from the dataset that is leveled with one. And after processing these data out model detect conspiracy successfully from 286 number of mail. And it predict 164 number of data as wrong detection.

Thus the accuracy of the model in the Red data set is = $286 * \frac{100}{450} = 65$

So the accuracy for the Red data is 65%.

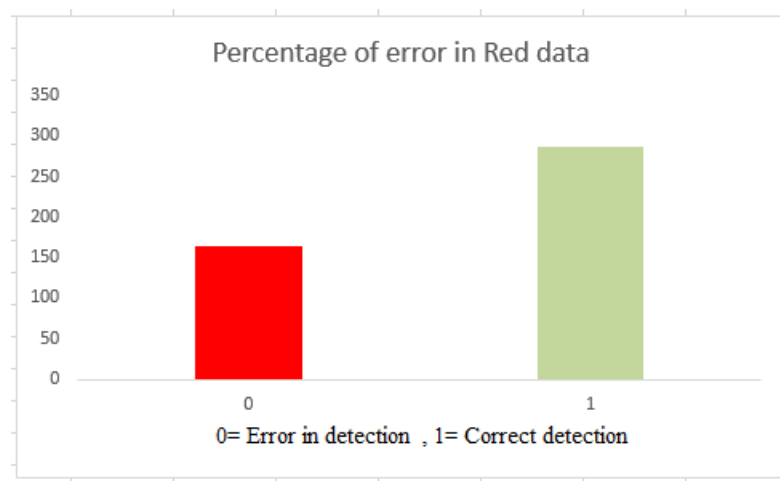


Fig 5.4: Percentage of error in Red Data

5.2.2 Evaluates from the real time mail

In the other way we can check the accuracy of our system in real time. For this way we can predict the accuracy of the system for real time email. In here we tested 100 mixed data send from one account to another account and count the number in four different ways. They are

1. True positive
2. True negative
3. False positive
4. False negative

True positive means the right detection of a Green email, true negative means predict a green mail as a red email, false positive means detect the Red data accurately, and false negative means incorrect detection of the Red data.

Here we are giving the table of this ratio for detection by the module in real time. Table 5.1 gives us the proper understanding of that concept

Total Email	True Positive	True negative	False positive	False negative
100	42	18	26	14

Table 5.1: Real time accuracy of mail data

So the overall real time accuracy of my system is $=(42 + 26) * \frac{100}{100} = 68$

Now we can say the real time accuracy of this system is 68%

So the overall accuracy of this system can be found by merging both the real time and the train set data is $= (324 + 286 + 42 + 26) * \frac{100}{(450+450+100)} = 67.8$

Thus we can say the accuracy is 67.8% overall.