A methodology for E-mail based detection of employee behavior change

Accepted 7th March, 2018

ABSTRACT

Organizational mail systems are an excellent source for detecting changes in employee behavior. However, this method is limited by privacy regulations. Alternatively, employees may decide to exploit their firm's organizational infrastructure by searching for an alternative workplace while on the job. The current paper demonstrates a method that utilizes Brownian motion to detect abnormalities in employee behavior using analysis of e-mail traffic meta data. This method is demonstrated using a well-known data set. In 73% of cases a change in e-mail patterns were observed six to ten weeks ahead of time for employees who had decided to leave their organization. The study also explains that the possible damage of false positive events in this case is minimal, thus, lending robustness to the methodology.

Keywords: Abnormality detection, e-mail traffic, random walk.

INTRODUCTION

Organizational damage caused by unplanned replacement of employees is estimated to be significant, especially in high-tech organizations where some employees may possess proprietary knowledge. Sexton et al. (2005) estimated the costs incurred by the unplanned replacement of employees due to loss of information, as 75 billion USD per year in the USA.

A common situation is that an employee may decide to exploit the organizational infrastructure, using the firm's mail system to search for alternative job offers, while still working for the organization. Such cases usually end with an employee giving a short notice to the organization about job termination. Sometimes this occurs after a long period of "fishing", during which the employee was "physically" a part of the organization, but had already "mentally" left her/his job.

Given the aforementioned phenomenon, one may assume that e-mails are an excellent source for mining information about the organizational social life. Tang et al. (2013) characterized several activities comprising e-mail mining. These include content inspection, discovery of human and organizational relationships, spam detection, E-mail categorization and property analysis.

Methods for e-mail mining always involve some form of text mining analysis. Several approaches have been used in the past to identify the behavioral changes of dissatisfied employees/customers using e-mail systems or message analysis. Most of this work is based on text mining with some use of clustering mechanisms, features selection or PCA. Holton (2009) tried to identify dissatisfied employees using Brill’s (1994) model. Spärck and Jones (1972) also identified disgruntled workers using text mining and searching for specific key words. Sakurai and Suyama (2005) tried to identify dissatisfied customers using a predefined dictionary and through text analysis. El-Alfy and Abdel-Aal (2011), in an attempt to identify spam mail, used the Group Method of Data Handling. Appavu et al. (2009) introduced an ID3-based algorithm for identifying mail containing threats based on a pre-defined dictionary. Gupta and Lehal (2009) described the main current techniques for the text mining of unstructured data. Kristof and Van den Poel (2008) suggested a method...
for prediction of customers churn based on e-mail.

Most of the aforementioned techniques are based on association rules, expert dictionaries or expert classification. The main problems with these approaches are privacy violation issues. In order to analyze and identify abnormal text messages, all of the aforementioned methods require actual reading of the text of the e-mail messages, a clear violation of privacy.

However, due to privacy regulations, employers’ ability to monitor the contents of e-mail messages is very limited, permitted only in special cases when possible crimes may have been committed. Given this limitation a method is required that may detect changes in e-mail patterns without reading content of the e-mail itself.

The current paper suggests an approach which differs from the majority of the work earlier described. This approach is based on measuring dynamic changes in e-mail metadata data indicators without reading the messages themselves. The purpose of this method is to identify changes in these indicators. It is worth noting that changes in these indicators may have various causes, for example, an employee’s unexpected health problems, financial problems or some other type of personal problem. Alternatively, a change in an employee’s position may often be followed by changes in her/his e-mail patterns. However, as explained, the benefits of lowering the number of false negatives, in this case, outweigh the disadvantages that may result from false positives. Hence, the misclassification of abnormal mail as indicating an employee’s intention to leave the company is less critical than the risk of being surprised by the sudden quitting of a team member who possesses critical knowledge.

The suggested approach is based on using a technique which simulates a random walk motion (RW) framework. The RW is used to detect changes in meta-data, such as the amount and density of e-mails sent and received inside and outside the organization, the timing and size of the email and the delay time before response, etc. RW is applied to measurements related to e-mail metadata without reading the actual message contents.

The analysis was applied to the Enron e-mail Dataset (2002), a well-known data set comprising 600,000 e-mail messages. The main result was as follows: 87 mail boxes from those of 158 employees contained daily mail messages sufficient for the analysis. Based on the alert time before the employee actually quit his or her job, a definition was made of false positive, true positive and false negative results. The performance of the algorithm was analyzed based on this definition. Based on this analysis, it was shown that in approximately 30 to 40% of cases the algorithm indicated changes in employee behavior 30 to 45 days before he/she actually left the organization.

In contrast to the situation in which an employee leaves his / her job is a surprise to the firm; using the current model, it was possible to detect significant numbers of cases ahead of time.

As previously indicated, it is worth noting that the presented method does not presume to predict an employee’s intention to leave their job. Our method simply hints that the employee is undergoing some type of change, which in some cases may be due to an intention to leave their job. However, it is suggested that such a change will usually be investigated by the employer. In most cases, the cost of a false positive in such a situation (if handled properly) may even have a positive effect since the employee will be flattered by the fact the the employer cares about him. On the other hand, the cost of a false negative is very high, as earlier explained.

### Random walk and the detection of changes: A dynamic process

The current paper uses an approach similar to those presented by Cheng (2009), Brill (2015) and Brill and Brill (2018). It measures the norm between two vectors of properties. In our case the properties are based on the metadata of the e-mail system.

Let us denote with vector $X_m$ the meta data which characterizes daily e-mail activity. The index $m$ denotes the day, while $X$ may have several measurements. The number of daily measurements is denoted by index $k$. Such measurements may include the number of e-mails sent or received, the average time delay between e-mails and the time of the day during which most e-mails are sent etc.

The value of $X_m$ can be normalized with the following process equation:

$$
\hat{v}_m^k = \frac{X_m^k - X_{\text{min}}^k}{X_{\text{max}}^k - X_{\text{min}}^k}
$$

(1)

Where $v_m^k$ is the normalized $k$ dimension of vector $V$ and $m$ is the discrete time index (in our case, days).

Given that $v_m^k$ is a normalized vector, the distance between two days can be measured by norms and is denoted by $B_m^h$. This measurement is calculated as the normalized Euclidian distance between two values of $V_m$ with a distance of $h$ days between them. It is given by Equation (2):

$$
B_m^h = \hat{V}_m - \hat{V}_{m-h}.
$$

(2)
Where $h$ is the difference in days between the two vectors. Figure 1 shows an illustration of this distance in a normalized two-dimensional space. In this case, $\hat{V}_m$ is a two-dimensional vector, that is, $k = 2$.

In terms of Figure 1, we assume a dataset with $M$ records and two variables in each record ($x_1$ and $x_2$). One may look at this dataset as a description of a virtual particle at each time stamp. After normalizing the dataset according to Equation 1, the Euclidian distance between each two points is the distance this particle travels. In the case where the distance is measured in a five-step gap, the result may be a chart as shown in Figure 1. Figure 2 shows a schematic chart of $B$ over time. Note that for the purposes of this paper, the distance used was seven records, that is, the distance between any two e-mails meta records over one week.

**IMPLEMENTATION OF THE RW APPROACH IN THE MAIL DOMAIN**

In order to detect abnormal behavior of an organization member using the e-mail system through the aforementioned framework, the Enron E-mail Dataset (2002) comprising more than half a million records was used. Using the daily mail log, a set of variables, as shown in Table 1 was calculated for each mail box. Each day is represented by a single line in the corresponding mailbox table. Each table (denoted by MCT —Mail Characteristics Table) represents the behavior of a single employee over the entire time period she/he was working for the firm. Such a table has $M$ lines (one line for each day) and $k$ columns (one column for each variable, as listed in Table 1).

Table 1 is only a suggestion, which was applied in the current research. One may choose other variables from the large collection of metadata related to the mail system.

Using the MCT of each employee, a vector of Bs was constructed for each mailbox. Assuming that the MCT contains $M$ lines, the B vector will have $M-1$ values. This vector is a set of $B(s)$ for each time stamp. The analysis of abnormal behavior is related to this vector.

Figure 3 shows a graphical illustration of the application of the algorithm over a single MCT. The illustration shows that the graph area can be divided into several segments. The first segment is from record index 0 to approximately record index 15 on the horizontal axis. These records represent the initial period of time when the employee began working in the organization. The second segment...
Table 1: Variables (columns) calculated based on mail system meta-data.

<table>
<thead>
<tr>
<th>S/No</th>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SEND_MAILS_OUT</td>
<td>Number of messages sent outside the organization</td>
</tr>
<tr>
<td>2</td>
<td>SEND_MAILS_IN</td>
<td>Number of messages sent inside the organization</td>
</tr>
<tr>
<td>3</td>
<td>TOTAL_SEND_MAILS</td>
<td>Total number of sent messages</td>
</tr>
<tr>
<td>4</td>
<td>AVG_BODY</td>
<td>Average length of message body in characters</td>
</tr>
<tr>
<td>5</td>
<td>STD_BODY</td>
<td>Standard deviation of (AVG_BODY)</td>
</tr>
<tr>
<td>6</td>
<td>SEND_HOUR_DIST1</td>
<td>Number of messages sent before the mid-day break</td>
</tr>
<tr>
<td>7</td>
<td>SEND_HOUR_DIST2</td>
<td>Number of messages sent after the mid-day break</td>
</tr>
<tr>
<td>8</td>
<td>SEND_HOUR_DIST3</td>
<td>Number of messages sent after normal work hours</td>
</tr>
<tr>
<td>9</td>
<td>RECEIVE_MAIL_FROM_OUT</td>
<td>Number of messages accepted outside the organization</td>
</tr>
<tr>
<td>10</td>
<td>RECEIVE_MAIL_FROM_IN</td>
<td>Number of messages accepted inside the organization</td>
</tr>
<tr>
<td>11</td>
<td>RECEIVE_MAIL_TOTAL</td>
<td>Total number of accepted messages</td>
</tr>
</tbody>
</table>

(Figure 3) is from record index 15 to 106; the third is from 106 to 180 and the fourth is the rest of the graph. Each of these segments has its own measurable characteristics, such as minimum, maximum, average, area under the graph and standard deviation of a moving window for \( W \) records (Where \( W \) is a configurable parameter).

Each segment describes a different behavior of the employee in the organization. It is assumed that the initial point of the graph is the \( k \) day, where e-mail traffic was first observed in the employee’s mailbox. This can also be considered as the employee’s first day + \( k \) in the organization. The last point in this graph is the last time e-mail traffic observed in the employee’s mailbox and can be also assumed to be the employee’s last day in the organization or within a few days prior to their last day.

The first and last segments are less interesting in this graph, as the first period is too short and the last segment is probably the period after the employee had already announced her/his intention to leave the organization since the graph values are very low, indicating small changes in the employee’s mailbox. In this case, the main focus would be the two other segments, which represent time periods prior to the end of employment—the second and the third segments. The idea is to look for a significant change in the graph. As an example, point 106 in Figure 3 on the horizontal axis is where the graph can be divided into two different behaviors. The fluctuations, minimum and maximum values and average values before and after this point on graph are different. Most values in the second segment are around 0.5 and vary from 0 to 1, at the most.
The third segment has different values. Here, the maximum values and minimum values are different, as the average and other measurable characteristics. A comparison between the characteristics of the two segments shows the following: the average $B$ value in the 2nd segment is 0.38, while the average $B$ value in the 3rd segment is 0.65. The standard deviations are 0.25 and 0.49, respectively, and the areas under the graph are 40.5 and 59.59 for the two segments. Using a simple t test to compare the normalized average of the second period with the normalized average of the third period (similar to the description in Equation 6 in the previous section) it is clear that the second and third regions are different.

By adding dates to the horizontal axis, it can be said that the employee started working in the organization at the end of June 1999 and finished working at the end of March, 2001; the point of change occurred somewhere around December, 2000 - one-and-a-half years after the beginning of her/his employment and four months before the end of the employment period. Thus, it can be assumed that at the time when a change is observed in the graph, the employee had decided to quit her/his job and maybe started looking for a new job. During this period, something in the employee's behavior changed and this was also expressed in the employee's e-mail traffic. It is unlikely that the employee was fired due to the time interval gap from the noticed change to the end of the data being too long to suggest a case of firing, as changes resulting from firing usually subside after a month or two. In the case of an employee being fired, we would expect to see a change at a different point in time, probably closer to the end of the employment period.

Figure 4A to d show four different examples of BM. Figure 4A and B show a significant change,
while Figure 4C shows no change and Figure 4D shows a change that is not significant.

Using the aforementioned approach, an analysis was performed for each of the 84 MCT(s). Each MCT was used as a detection tool to identify possible indications that the employee is about to quit her/his job. Within this analysis process, the following definition was attached to each MCT: Two check points (A and B) as shown in Figure 5 were set according to the following criteria:

In cases where a change in the chart of the MCT was before point A, the alert of the algorithm was declared as a false positive (FP), that is, the change in the BM line was not an indication of an intention to leave the organization.

In cases where a change in the chart of the MCT was after point A and before point B the alert of the algorithm was declared as a true positive (TP), that is, the change in the BM line was an indication of an intention of the employee to leave the organization.

Finally, in cases where the alert of the algorithm was declared after point B and before job termination, the alert was declared as false negative (FN), that is, the alert was given too late and the algorithm missed the intention of the employee to leave the organization. The values A and B were set to different values as shown in Table 2. Setting the distance between FP and TP to 75 and the distance between TP and FN at 30 yielded 29 FP, 25 TP and 37 FN results. Changing the distances to 90 and 15, respectively, yielded 29 FP, 47 TP and 21 FN.

**CONCLUSION**

The current paper demonstrates the use of a random walk
methodology to detect abnormal behavior in employees using an analysis of e-mail traffic metadata. The detection is achieved by analyzing e-mail traffic metadata, which is public data and can be inspected by any mail provider without violating the users' privacy. The results of the analysis of known e-mail datasets showed that in some companies, 48% of employee resignations could have been predicted by such an algorithm, in a time window comprising 15 to 90 days before the employee actually terminated their employment. It is also noted that the algorithm may use different variables as long as it captures the e-mail traffic variability.

Unlike other methods earlier discussed, which require interpretation of the text by a human expert or the use of a defined dictionary with human classification, the aforementioned method does not depend on human expertise. It can be performed by anyone capable of analyzing an e-mail system log file.

Finally, the current method is less vulnerable to damage caused by false positive results. This is due to the fact that an employer is always expected to be interested in the well-being of his/her employees.

REFERENCES


