Expert Systems With Applications 135 (2019) 351–361

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Employee profiling via aspect-based sentiment and network for insider threats detection

Charlie Soh*, Sicheng Yu, Annamalai Narayanan, Santhiya Duraisamy, Lihui Chen

Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 31 December 2018
Revised 14 May 2019
Accepted 27 May 2019
Available online 28 May 2019

Keywords:
Insider threat
Aspect-based sentiment analysis
Network embedding
Anomaly detection

ABSTRACT

Historically, the harm caused by insiders has proven to be one of the greatest concerns for any organization. As such, it has received considerable attention from both the industrial and research communities. Existing works mainly focused on modeling the employees' normal biometric behavior (e.g., human to device interaction pattern) to detect anomalous behavior which corresponds to the insider activity. However, it is unattainable to stop the insider at the final moment when the malicious act is being carried out. In this paper, we propose a novel framework which performs employee profiling based on aspect-based sentiments and social network information and examine its applicability for early detection of potential insider threats. On the contrary to the traditional sentiment analysis, aspect-based sentiment analysis provides more fine-grained information on the employee. Our framework employs a combination of deep learning techniques such as Gated Recurrent Unit (GRU) and skipgram to build temporal sentiment profiles for the employees. It then performs anomaly detection on the profiles and ranks the employees based on their respective anomaly score. Due to the absence of relevant benchmark dataset, we augmented the publicly available real-world *Enron* email corpus with an insider threat scenario to evaluate our framework. The evaluation results demonstrate that the augmentation is indeed reflected in the augmented employee's anomaly ranking (i.e., from normal to abnormal) and her close associates are indeed placed closely to her when the profiles are visualized in the 2D space. The profiles obtained from our framework can also be used to complement any existing expert and intelligent systems with additional capabilities in handling textual information such as, integration with profiles obtained from biometric behavior to form a more comprehensive threat detection system.

1. Introduction

Today, insider threat has become one of the major concerns for organizations. Insiders are people with authorized access to sensitive information in an organization. The trust afforded to employees, while necessary for them to perform their tasks, exposes the organization to a wide range of insider attacks. It was reported that the damage an insider could have dealt to an organization is far worse compared to outsider attacks and could cost as much as $26.5 million[1]. Despite the extensive effort from both the industrial and research communities to combat the threats, there is a rising trend in all variations of insider threats[2]. In particular, the cases of sabotage insider attacks, such as the recent Tesla case,[3] have increased by over 60% over the past two years (Ponemon, 2011).

A large majority of the existing defense solutions focus on modeling the employees' normal biometric behavior (e.g., mouse and keyboard usage) and/or network logs to detect anomalous behavior (Liu, De Vel, Han, Zhang, & Xiang, 2018). However, typical organization has a complex infrastructure and is comprised of a mix of people from different backgrounds, where each of them may have a different role in the organization. For these reasons, the insider threat problem is considerably more elusive than any other threats that the organization faces and cannot be addressed by technological means alone.

* Corresponding author.
E-mail addresses: csoh004@e.ntu.edu.sg (C. Soh), yu0021ng@e.ntu.edu.sg (S. Yu), annamala002@e.ntu.edu.sg (A. Narayanan), santhiya003@e.ntu.edu.sg (S. Duraisamy), elhchen@ntu.edu.sg (L. Chen).

[1] https://globenewswire.com/news-release/2018/10/22/1624809/0/en/Cybercrime-Continues-to-Plague-Organizations-as-Concerns-over-Security-Breaches-Increase.html.

[2] https://www.ca.com/content/dam/ca/us/files/ebook/insider-threat-report.pdf.

[3] https://www.darkreading.com/the-6-worst-insider-attacks-of-2018---so-far/d/d-id/1332183.

https://doi.org/10.1016/j.eswa.2019.05.043
0957-4174/© 2019 Elsevier Ltd. All rights reserved.

Furthermore, it is nearly impossible to stop the insider right before the incident, hence, the best solution is to prevent the threats from occurring through early detection (i.e., observable signs that precede the incident). A recent study (Homoliak, Toffalini, Guarnizo, Elovici, & Ochoa, 2018) has observed that insiders typically experience a phase of emotional changes before the insider commits the malicious act. Reinforcing this fact, a previous study (Ponemon, 2011) has reported that this phase of emotional change is on an average of 73 days. Moreover, case studies documented in CERT (Kowalski et al., 2008) advocate that insiders initiate atypical communications to others. These findings suggest that emotional/ psychological and social factors observed in social communication channels are critical in addressing early detection of insider threats.

However, developing a system to perform such task poses a number of challenges. Firstly, due to the sheer volume of communications that an organization may experience everyday, it is impractical to manually vet through all the communication channels and note the transformation of the employees. Secondly, the first challenge can be addressed through the use of machine learning algorithms, which typically require high quality samples from both positive (i.e., communication content associated with insider threats) and negative (i.e., normal communication content) classes to train. However, such data are very scarce not only due to privacy concerns but also because revealing of the information on such incidents could harm the reputation of the organization. Furthermore, the scarcity of real-world data also hinders the assessing of defense solutions. Lastly, not all communication contents are essential and not all emotional changes are relevant for insider threat analysis. In fact, most of the contents are irrelevant and considering them in the analysis may dilute the features and affect the model's performance.

To this end, we propose a profiling framework called ASEP, which stands for Aspect-based Sentiment Employee Profiling. ASEP builds temporal profiles which incorporate sentiment and network information of the employees. More specifically, inspired by the recent success in aspect-based sentiment analysis (ABSA), we propose to model the emotional/ psychological state of the employees by performing ABSA on their email contents (see §3.2). Al Tabash and Happa (2018) demonstrate that capitalizing on expert knowledge to furnish external information that is obscure to the detection model can significantly improve the detection accuracy. In this work, it is tough for the detection system to automatically identify which are the important or relevant emotional/opinion changes for user profiling. Therefore, in our system, expert knowledge can be accommodated via the relevant aspects, which can be specified by an organization based on the characteristics of the organization. In this way, only opinions/emotion with respect to those relevant aspects will be taken into consideration for user profiling. After performing ABSA, our network embedding model combines ABSA results and social network information to obtain the sentiment profiles (see §3.3). In addition, to address the challenge on the deficit of dataset, we augment the publicly available *Enron* dataset (Klimt & Yang, 2004) by transforming an employee into an insider based on a simulated insider scenario with reference to those described in CERT (Cappelli, Moore, & Trzeciak, 2012) (see §4.1) and use it to evaluate our framework. The augmented dataset is made publicly available[4].

In summary, we make the following contributions in this paper:

- We present a novel employee profiling framework with deep learning models for insider threat detection based on aspect-based sentiment and social network information.
- Even though benchmark dataset for insider threats detection are available, they are not suitable for evaluating our proposed framework. Therefore, we simulated an insider scenario in the real-world *Enron* email corpus. To encourage further research in this direction, we make the augmented dataset publicly available.[4]
- We evaluate ASEP on the augmented *Enron* dataset and demonstrate that the augmentation is successfully captured by the employee profiles and reflected in the anomaly ranking.
- We visualize the profiles and demonstrate that the profiles also capture the network properties, such that close associates are placed closely in the 2D space.

2. Related work

A plethora of studies has been conducted in an attempt to address the insider threats. Hence, over time multiple taxonomies have been suggested to categorize these studies and the insider attacks. Based on the analysis of various real-world case studies, CERT (Cappelli et al., 2012) suggests that the insider attacks can be categorized into three main types, namely, IT sabotage, IP theft, and insider fraud. A recent survey (Homoliak et al., 2018), suggests that these studies can be generally categorized as, incidents and datasets, analysis of attackers, simulations and defense solutions. In particular, this work aims to propose a defense solution and more specifically, a framework for the detection and assessment of insider threats. The existing studies having the same objective typically view it as an anomaly detection task and the solutions usually involve the monitoring of observables such as network, host, contextual and biological data (Homoliak et al., 2018).

Network data typically refers to network logs or network traffic information or even both. For example, Liu et al. (2009) propose a framework to perform statistical analysis on network traffic for detecting exfiltration of sensitive data. Sibai and Menascé (2011) propose a framework, Autonomic Violation Prevention System (AVPS) which enhance traditional Network Intrusion Prevention Systems (NIPS) with the ability to protect the organization from security policy violations even with legitimate access. Pagliari et al. (2015) propose an approach to perform bi-clustering and one-class SVM on weak network attack indicators. As a result, the approach provides the capability to determine the cause of the detected anomaly. Alotibi, Clarke, Li, and Furnell (2016) build biometric-based behavioral profile, from raw network traffic metadata to identify user's application-level interactions. The resulting profiles can then be used to detect insider threats.

Several other studies instead focus on host data, which largely deals with data produced from human interactions with the host device (i.e., computer) such as system calls, keyboard strokes, mouse clicks, etc. For example, Song, Salem, Hershkop, and Stolfo (2013) profile users based on system level events, such as process creation, registry key changes, and file system actions. The profiles encompasses the users' biometric behaviors and can be used for masquerade and insider threat detection. While not focusing on providing a defense solution to insider threats, Camiña, Hernández-Gracidas, Monroy, and Trejo (2014) proposed a synthetic dataset namely, the Windows-Users and -Intruder simulations Logs (WUIL) dataset, which encompasses faithful masquerade attempts, in hope to motivate the research community to use WUIL and further explore the solution to masquerade detection. Leu, Tsai, Hsiao, and Yang (2017) propose an Internal Intrusion Detection and Protection System (IIDPS) to detect insider attacks at the system call level. The system uses data mining and forensic techniques to keep track of users' usage habits as their forensic features and compare the current usage pattern with the existing profile to determine if the current user is valid.

[4] https://sites.google.com/view/enronplus/home.

These network and host based approaches focus on identifying the anomalous human to device interaction pattern that precedes the malicious act. However, such approaches may suffer from high false positives due to the inherently noisy data and the anomalous pattern may only arise during the act and discovered after the damage is done.

To avoid the limitations and also complement the above mentioned approaches some recent studies shift their attention to utilize contextual data. These studies can generally be classified into psychology and sentiment analysis, both of which are typically based on linguistics. The psychology studies mainly use Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010) to identify the psychological characteristics of a potential insider. For example, Brown, Watkins, and Greitzer (2013) perform an investigation in the potential for monitoring electronic communications, such as emails, for clues to identify potential insider threats. They conduct an experiment with *Enron* email corpus and LIWC dictionaries and the results suggest that there is a subtle but measurable contrast that may provide clues. Rather than a detection framework, Legg, Buckley, Goldsmith, and Creese (2014) propose a visual analytics system to aid the assessment of socio-linguistic behaviors in emails. While the system allows customizable decision making, it demands expertise and feature engineering effort from the analyst. Nevertheless, the system can be adapted to complement the detection approaches and allow the results to be visualized and explained. Alahmadi, Legg, and Nurse (2015) investigates the relationship between internet browsing content and the OCEAN personality characterization to identify potential insider threats. Chi, Scarllet, Prodanoff, and Hubbard (2016) combine linguistic analysis and K-means algorithm for analyzing communications such as emails, to determine whether the employee meets certain personality criteria and compute a risk level for each employee. On the other hand, sentiment analysis focuses more on the emotional state of an individual. Such as whether the individual is feeling positive, negative or neutral. Potential insider threats are typically associated with negative sentiments. For example, Park, You, and Lee (2018) perform sentiment analysis on social media (e.g., tweets) to determine the users' sentiment scores and negative tweets ratios, then classify the users by threat level. However, psychological characteristics are general and are not indicative of potential insider threats by themselves. Similarly, the traditional sentiment analysis may be to coarse-gained for accurate detection. To address such limitations, ASEP also monitors electronic communications, but focuses on detecting anomalous aspect-based sentiment (a more fine-grained sentiment analysis) rather than personality traits.

Diversely, with the advance in bioscience, several studies which explore the use of biological behaviors for insider threat monitoring and detection have emerged. These studies use biological signals such as electrocardiogram, skin temperature, voice change etc., to monitor the physiological or emotional change in the potential insider. For example, Almehmadi and El-Khatib (2014) proposed a Physiological Signals Monitoring system which is capable of detecting the insider threat through the change mainly in the insider's electrocardiogram, skin temperature and Galvanic Skin Response before the threat is being executed. Although promising, these approaches may have limited practical use due to legal and economic concerns.

3. Methodology

An overview of ASEP is presented in Fig. 1. As depicted, our framework consists of four main modules, namely, preprocessing, ABSA, employee profiling, and anomaly detection. The main processes of the framework are summarized in Algorithm 1.

Algorithm 1 ASEP overview.

```
Input:
    Email data and list of aspects
Output:
    Employee profiles and anomaly ranking.

    Initialization: Pre-trained ABSA module, randomly initialize
    skipgram parameters $\Phi^{skipgram}$
    for each iteration t do
        Preprocess the raw email data to extract aspect based
          sentences and collect the network information like
          sender and receiver IDs.
        Apply ABSA with recurrent attention to get aspect
          level sentiments.
        Update $\Phi_t^{skipgram}$ based on $\Phi_{t-1}^{skipgram}$, ABSA output and
          network information to learn employee profiles.
        Utilize isolation forest to compute anomaly score
          from the profiles and rank them accordingly.
    end for
```

3.1. Preprocessing

In the preprocessing module, the incoming emails will be parsed to extract aspect-based sentences along with the network information. First and foremost, aspects relevant to the organization and insider threat detection need to be identified. If the email corpus is available (i.e., email communications over a certain period has already been collected), several unsupervised aspect extraction techniques can be applied (Bagheri, Saraee, & de Jong, 2013; Qiu, Liu, Bu, & Chen, 2011; Wu, Zhang, Huang, & Wu, 2009) to obtained a list of aspect to choose from, otherwise, they can be manually chosen. Since our objective is to build a profile for each employee, for each email thread (i.e., email messages that are related by a response such as forward or reply) we only consider the most recent email. Furthermore, each email may encompass several sentences and thus, may have multiple aspects with different sentiments which increases the complexity of the sentiment analysis (Chen, Xu, He, & Wang, 2017). Therefore, to better capture the aspect-based sentiment, we first decompose the emails into individual aspect-based sentences. Then for each sentence, if it contains at least one of the selected aspects, we consider it as an aspect-based sentence and include it in our further analysis, otherwise, the sentence will be ignored. In cases where the sentence consists of multiple aspects, we treat it as multiple aspect-based sentences, where each sentence is paired with each respective aspect while ignoring the others. Here we also extract the timestamp and network information from the email header.

3.2. Aspect-based sentiment analysis

The main objective of the ABSA module is to perform sentiment analysis towards a targeted aspect and classify the sentiment of each aspect-based sentence in the emails with respect to the corresponding aspect word. The detailed architecture of our ABSA module is as shown in Fig. 2 and its process are summarized in Algorithm 2. It encompasses four main layers.

Let the words of the input sentence S be $\left[s_1, s_2, .., s_N\right]$, and aspect term T be $\left[t_1, t_2, .., t_M\right]$, where N and M represent the length of the sentence and the length of aspect term, respectively. We concatenate pre-trained word vector from GloVe (Pennington, Socher, & Manning, 2014) and AffectiveSpace (Cambria, Fu, Bisio, & Poria, 2015) to build the vector $s_n \in \mathbb{R}^d$ for each word. Where the former captures the fine-grained semantic and syntactic regularities

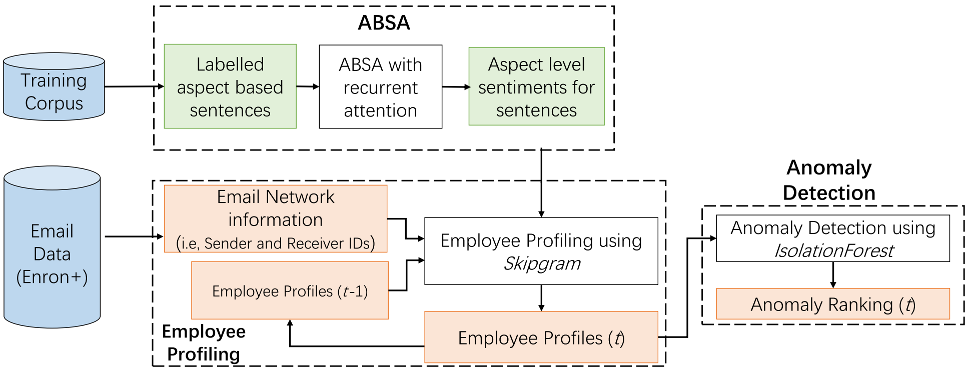

ABSA
Training Corpus
Labelled aspect based sentences
ABSA with recurrent attention
Aspect level sentiments for sentences
Anomaly Detection
Email Data (Enron+)
Email Network information (i.e, Sender and Receiver IDs)
Employee Profiling using Skipgram
Anomaly Detection using IsolationForest
Employee Profiles (t-1)
Anomaly Ranking (t)
Employee Profiling
Employee Profiles (t)

Fig. 1. ASEP overview.

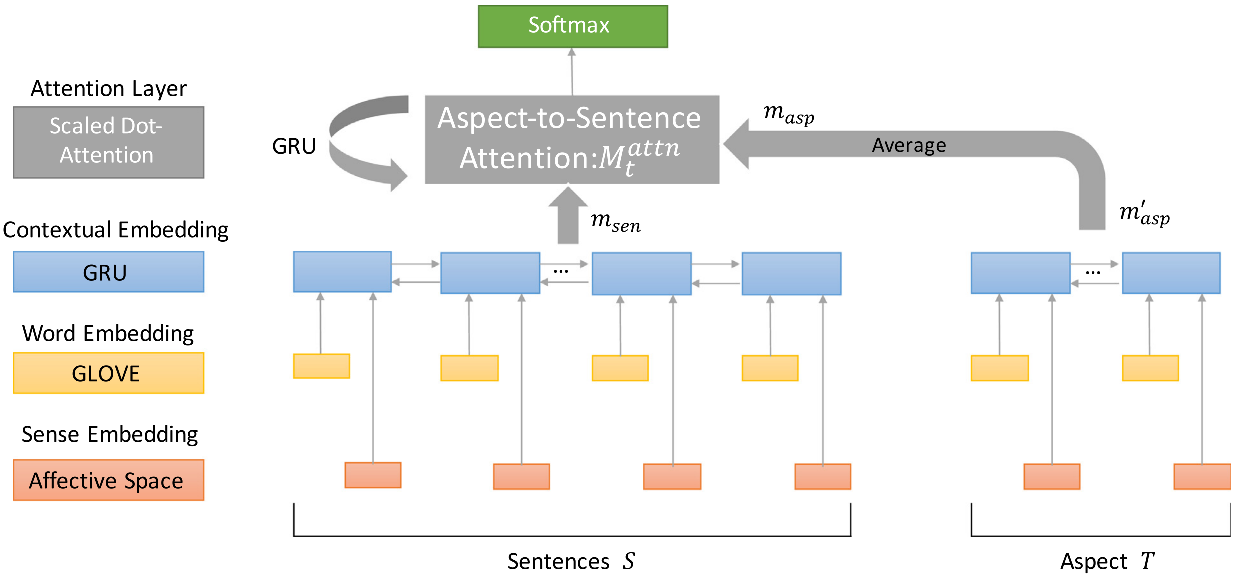


Fig. 2. ABSA module.

while the latter incorporates common sense and affective knowledge, thus helps to improve the performance of our ABSA model.

We use two separate bi-direction GRUs (Cho, Van Merriënboer, Bahdanau, & Bengio, 2014) to learn the representation for the input sentence and the corresponding aspect term, respectively. Both classes of gated units, LSTM (Hochreiter & Schmidhuber, 1997) and GRU, can address the vanishing gradient problem to a certain extent. While both units had achieved similar performance (Chung, Gulcehre, Cho, & Bengio, 2014), we choose GRU as it has fewer parameters to optimize compared to LSTM. GRU updates the hidden state at each time step by considering the hidden state at the previous time step and the input of the current time step. The update process of GRU is as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_{t-1}])$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_{t-1}])$$

$$\tilde{h_t} = tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h_t}$$

$$y_t = \sigma(W_o \cdot h_t)$$

Hence, the representation of the input sentence and aspect term is modeled as:

$$m_{sen} = (\overrightarrow{GRU_{sen}}(S), \overleftarrow{GRU_{sen}}(S))$$

$$\acute{m}_{asp} = (\overrightarrow{GRU_{asp}}(T), \overleftarrow{GRU_{asp}}(T))$$

In cases where the aspect term encompasses more than one word, we use the mean of the aspect term's representation obtained at each time step to represent the aspect term in our subsequent computations.

$$m_{asp} = \frac{1}{M} \sum \acute{m}_{asp}$$

More importantly, the attention mechanism in our ABSA module allows it to learn the important parts of the input sentence that significantly influence the sentiment prediction. Sentences found in social communications, such as emails, are inherently more complex than review sentences. Attending to such complex sentences with single attention may fail to identify the important parts of the sentence. Hence, as inspired by Chen, Sun, Bing, and Yang (2017), we employ recurrent attention which we describe in the following. In iteration t, m_{sen} is weighted as follows:

$$M_t^{attn} = \sum_N \alpha_j^t \cdot m_{sen,j}$$

where $m_{sen,j}$ refers to the j^{th} term in m_{sen} and α_j^t is the attention weights of j^{th} word towards the corresponding aspect which

Algorithm 2 ABSA module.

```
Input:
    A set of input sentence S = [$s_1$, $s_2$,..., $s_N$] and aspect term T = [$t_1$, $t_2$,..., $t_M$], where N and M represents the length of sentence and term, respectively. $W_Q$ and $W_K$ as parameter matrices to be trained and $d$ the dimension of word vector.
Output:
    The label that classifies the aspect-based sentiment into one of the three sentiment classes, positive, negative or neutral.

    Initialization: Concatenate pre-trained word vector from GloVe and AffectiveSpace to build the vector $s_n \in \mathbb{R}^d$ for each word.
    for each iteration do
      for each time step in iteration $t$ do
        $m_{sen} = (\overrightarrow{GRU_{sen}}(S), \overleftarrow{GRU_{sen}}(S))$
        $m\acute{}_{asp} = (\overrightarrow{GRU_{asp}}(T), \overleftarrow{GRU_{asp}}(T))$
        $m_{asp} = \frac{1}{M}\sum m\acute{}_{asp}$
        $M_t^{attn} = \sum_N \alpha_j^t \cdot m_{sen,j}$
        $e_t = GRU(M_t^{attn}, e_{t-1})$
      end for
      $pred = softmax(e_t)$
    end for
```

is computed based on Scaled Dot-Product Attention (Vaswani et al., 2017) with one head as follows:

$$Attention(Q, K) = softmax\left(\frac{QW_Q \cdot KW_K}{\sqrt{d}}\right)$$

where W_Q and W_K are parameter matrices to be trained and d is two times the hidden dimension of GRU (bi-directional). After calculation of M_t^{attn} in iteration t, we use another GRU initialized with m_{asp} to combine it with the output at iteration $t-1$:

$$e_t = GRU(M_t^{attn}, e_{t-1})$$

At the last iteration, the output serves as the sentence embedding and goes through a *softmax* layer for classification of the aspect-based sentiment into one of the three sentiment classes, positive, negative or neutral.

3.3. Employee profiling

Given the ABSA predictions and network information, the employee profiling module generates temporal distributed representations for each employee (i.e., employee profile) in the organization. The employee profiles help organizations to identify the potential insiders whose aspect-based sentiments differ from others and enable the visualization of their social communities.

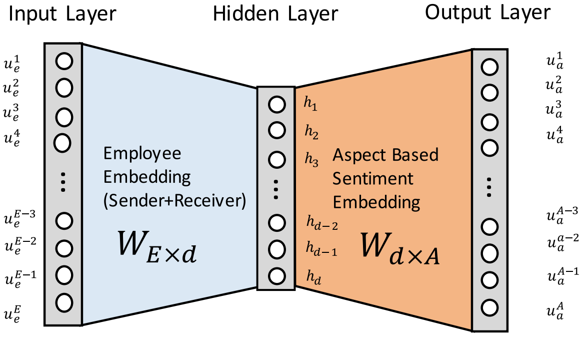


Fig. 3. Employee profiling model.

Our employee profiling module is based on the skipgram model(Mikolov, Chen, Corrado, & Dean, 2013) that has achieved excellent results in word/document (Le & Mikolov, 2014; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) and network (Grover & Leskovec, 2016; Narayanan, Chandramohan, Chen, Liu, & Saminathan, 2016; Perozzi, Al-Rfou, & Skiena, 2014) embeddings. More specifically, given a particular aspect-based sentence, the skipgram model learns the employee's d-dimension representation by predicting the most relevant aspect-based sentiment such that, the employee having negative sentiment towards the aspect *'business'* will have distinctive vector compared to the employee having neutral sentiment towards *'business'*. Therefore, as depicted in Fig. 3, our employee profiling skipgram model has input vectors which represent the employees and output vectors which represent the aspect-based sentiments. The objective of profiling module is to maximize the following average log probability:

$$\frac{1}{M}\sum_{i=1}^{M}\log p(u_a^i|u_e^i)$$

where M is the total number of aspect-based sentiment and associated employee pairs, while u_a^i and u_e^i denote the aspect-based sentiment and the employee (i.e., sender/receiver) corresponding to the i^{th} aspect-based sentence, respectively. In the above equation, we do not explicitly encode the network information into the model. However, as demonstrated in our evaluation, by assuming that the sender and the receiver share the same sentiment the model is able to capture correlation between the employees. The probability function $p(u_a^i|u_e^i)$ is computed as follows:

$$p(u_a|u_e) = \frac{exp(v_{u_a}^{\intercal} v_{u_e})}{\sum_{i=1}^{A} exp(v_i^{\intercal} v_{u_e})}$$

where v_{u_e} refers to the vector of a specific employee to be trained and v_{u_a} refers to the corresponding aspect-based sentiment, while A and E represent the total number of aspect-based sentiments and the total number of employees within this time interval, respectively.

Furthermore, the model is trained in an incremental fashion, in other words, the employee profiles obtained in the previous interval will be used as the initial embeddings in the next interval.

3.4. Anomaly detection

Lastly, given the temporal profiles the anomaly detection module performs anomaly detection based on the isolation forest algorithm (Liu, Ting, & Zhou, 2008). Isolation forest is an unsupervised method which randomly selects i number of samples to build n number of isolation trees (iTree) with a random selection of features and splits. As shown in Fig. 4, the intuition is that since there are fewer instances of anomalies and they have different attribute values from normal instances, therefore they will be more susceptible to isolation and are isolated closer to the root of the tree compared to the normal samples. Consequently, given the unique characteristic of isolation forest, it has demonstrated promising results on sparse and high dimensional data set with a large number of irrelevant attributes (Zhu, Zeng, & Kosorok, 2015).

More specifically, each iTree T is a binary tree in which every node has either no child or two children nodes (T_l, T_r). Given a dataset of i instances $X = \{x_1, \ldots, x_i\}$, a subset $X' \in X$ is used to build an iTree. X' is recursively divided by a randomly selected attribute a and a random split value v until either the node has only one instance or all data at the node have the same values. The path length of an instance x_i is measured by the counts of edges it traverses from the root node to the external node. Lastly, anomaly scores $s(x_i, n)$ are computed based on the computation of the

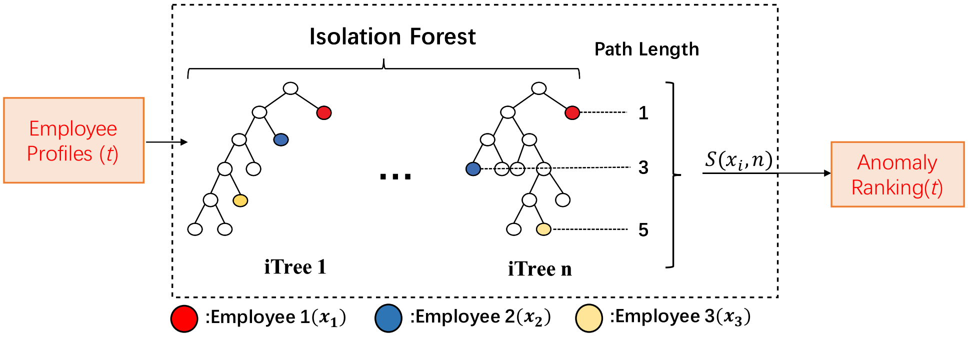


Fig. 4. Anomaly detection.

average path length of unsuccessful searches in Binary Search Tree (Liu, Ting et al., 2008).

In this framework, for each employee profile, the isolation forest algorithm will compute an anomaly score and we obtain the anomaly ranking of the employees by sorting these scores in ascending order.

4. Evaluation

In this section, we evaluate the performance of ASEP. We have implemented a prototype of ASEP in Python[5] programming language and the experiments are run on a server with IntelXeon(R) E5-2640 2.40GHz CPU and NVIDIA Tesla V100 GPU card running Ubuntu 16.04. In the rest of this section, we first discuss the dataset that we will be using in the evaluation. Then move on to describe the details of the experiment setup before we present and discuss the results.

4.1. Dataset

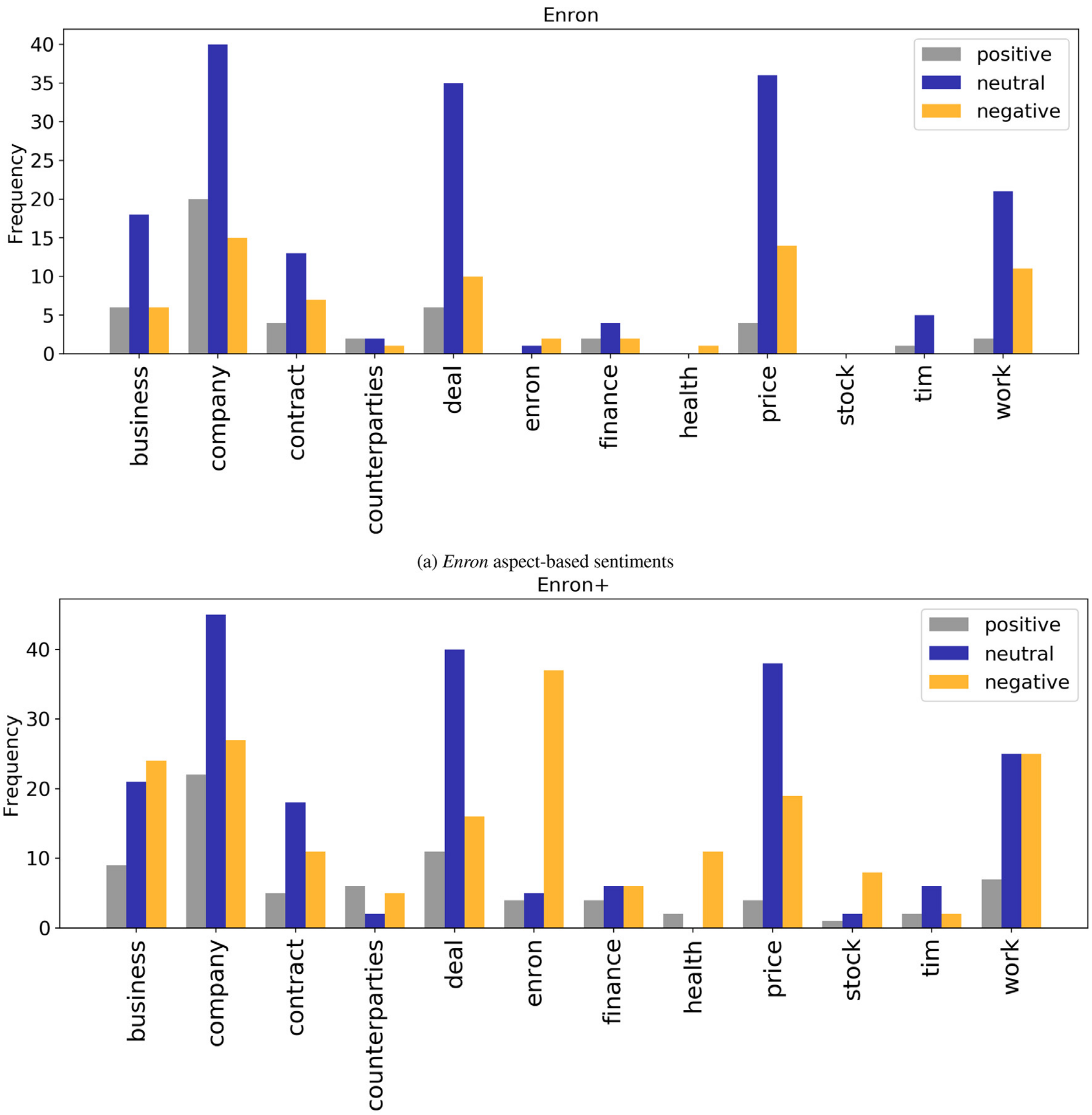
To the best of our knowledge, the *Enron* email corpus is the only publicly available real-world mass collection of organization email corpus. The reason for the scarcity of such dataset is mainly because the disclosure of such data is usually bounded by numerous privacy and legal restrictions and in cases where insider threats are involved, it may also affect the reputation of the organization. Interestingly, the executives of *Enron* employed accounting practices that falsely inflated its revenues, resulting in one of the biggest accounting fraud in history which led *Enron* to bankruptcy. Nevertheless, the corpus as it stands does not encompass an insider threat scenario that we aim to address in this work. One possible solution to dataset deficiency is to create a synthetic dataset, however, doing so would require a large amount of effort and the resulting corpus may not be realistic. Therefore, to evaluate the efficacy of this work, we choose to augment the *Enron* corpus.

More specifically, before we augment the *Enron* corpus, we first refer to the list of real-world case studies presented in the CERT guide to insider threats (Cappelli et al., 2012). Thereafter, we analyze the *Enron* corpus to identify a suitable candidate for augmentation to imitate a scenario in the case study. We discover an interesting situation from a study which uses *Enron* corpus as an example to analyze company structure from a social network (Palus, Bródka, & Kazienko, 2010). That is, while managers typically have far higher social score than others, a specialist in the Northwest has a higher social score than her manager. We further analyze the emails of said specialist using the NRC Emotion Lexicon (Mohammad & Turney, 2013) which is a collection of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). Interestingly, towards the end of December 2001 to mid of February 2002, when *Enron* was going through bankruptcy, her negative sentiment and emotions were relatively higher compared to the previous months. Hence, we read her emails and observe that before and after November 2001, her emotional fluctuations towards the company and her private life was evident from her emails. Therefore, with reference to the real-world case studies provided in CERT book, we augmented the *Enron* corpus and in particular the ego network of said specialist to include an insider scenario. We refer the augmented *Enron* corpus as *Enron+*. More specifically, in *Enron+* we transformed the said specialist into a potential insider by incorporating emails which encompass a plot that amplifies her emotions between September 2001 to February 2002. The simulated plot is to have the specialist experience a gradual emotional change and become especially negative in January and February 2002. Also, to evaluate the capabilities of our profile in representing community information, we augment the said specialist to have frequent communication with two others namely, $associate_A$ and $associate_B$. To contribute to the research community and encourage further research in this direction, the emails of the augmented ego network are made available[4].

There are several different versions of the *Enron* corpus available online. In particular, the *Enron* corpus used in this work is obtained from (Klimt & Yang, 2004), which consists of email data over a timeline from December 2000 to October 2002. Despite being a big organization with a large number of employees, the *Enron* corpus only contains email data from 150 employees and is further downsized due to redaction effort. Furthermore, since the insider scenario is simulated to occur between September 2001 to February 2002, we focus on these periods in this experiment. The emails for each employee are sorted into folders (i.e., *inbox*, *sent*, *deleted_items*, etc.). Given that our goal is to build a profile for each employee in the organization we only consider the *sent* and *sent_items* folder. Moreover, to reduce noisy data, we also excluded the employees who sent and received a total of fewer than 50 emails in the predefined time interval of a month.

As mentioned in Section 3.1, it is necessary to first identify the relevant aspects before we can extract the aspect-based sentences from the emails. However, different organization may have a different set of relevant aspect terms. For example, the aspect term ’enron’ and ’price’ are critical aspects for analyzing insider threats in the *Enron* organization, but may not be relevant to other organization. In this experiment, since the corpus is available, we first perform aspect extraction based on double propagation (DP) (Qiu et al., 2011) then manually select the relevant aspects from the top 50 aspects returned by DP. More specifically, we have selected a total of 12 aspects. The frequency distributions of the aspect-based sentiments for the transformed specialist in *Enron* and *Enron+* are presented in Fig. 5a and b, respectively.

[5] https://www.python.org/.



(a) *Enron* aspect-based sentiments

(b) *Enron+* aspect-based sentiments

Fig. 5. Frequency distribution of aspect-based sentiments for augmented specialist: *Enron* vs*Enron+*.

4.2. Experiment design

We compare the accuracy of our ABSA model to several other baseline models. More specifically, our ABSA model is based on a supervised learning model which requires known sentiment (i.e., positive, negative, neutral) labels for training. Hence, we manually labeled a subset of the *Enron+* corpus and use them to train the model. Furthermore, to enhance the generalizability of the model, we include the restaurant and laptop review datasets from SemEval 2014 Task 4 (Pontiki et al., 2014) in the training to augment the training data. However, since such information is not leveraged in the baseline approaches, we also evaluate our ABSA model without including restaurant and laptop review datasets in training the model to provide a fair comparison. Similar to the restaurant and laptop review datasets, 80% of the labeled email sentences are used for training and the remaining 20% are used to measure the accuracy of the models. The rationale for choosing 80/20 train to test set split ratio instead of other common split ratio such as 90/10 is due to the small size of the manually labeled data. The training is performed with a batch size of 128. We optimize the model using Adam (Kingma & Ba, 2014) with a learning rate of 0.001. The pre-trained word vectors are obtained from 300-dimension GloVe vectors (Pennington et al., 2014) and concatenated with 100-dimension AffectiveSpace vectors (Cambria et al., 2015) where available. For cases where a particular word is not found in either pre-trained word embeddings, the respective portion of the word vector will be initialized with zeros. Three iterations of recurrent attention are performed to obtain the final sentence embedding. The trained ABSA model is then used to perform aspect-based sentiment classification for all the extracted aspect-based sentences and the results are subsequently passed to the employee profiling module.

Due to the redaction effort on the *Enron* corpus, the volume of email communications per day is low. Therefore, in this experiment, we build the employee profiles at a monthly interval, where each profile is represented as a vector of 128-dimension. However, for real-world large organizations, profiling should be performed more frequently. Unlike the ABSA module, the employee profiling module does not perform supervised learning, but instead uses an unsupervised representation learning approach to build the

Table 1
ABSA baseline comparison results in Accuracy(%).

Dataset	Emails	Tweets
TD-LSTM (Tang, Qin, Feng, & Liu, 2016)	55.88	70.80
IAN (Ma, Li, Zhang, & Wang, 2017)	61.60	NA
MemNet (Tang, Qin, & Liu, 2016)	57.94	69.65
CABASC (Liu, Zhang et al., 2018)	57.23	71.53
Our ABSA (without restaurant/ laptop training data)	65.58	70.38
Our ABSA	68.44	71.97

profiles and therefore cannot be evaluated in a similar fashion as the ABSA module. Hence, we evaluate the quality of our employees profiles in terms of whether the employees who share similar aspect-based sentiment have similar profiles and whether the employee profiles capture the implicit social network information.

To examine whether employees who share similar aspect-based sentiment have similar profiles, we analyze the correlation between pairwise similarities in aspect-based sentiments and profiles. We first compute the pairwise similarities for all employees, in terms of their aspect-based sentiments. More specifically, the similarities are computed via dot product, $Simiarity(A, B) = V_A \cdot V_B^T$. Particularly, when computing aspect-based sentiments similarity, V_A and V_B are the vector representation of aspect-based sentiments (*freq profile*) for employee A and B, respectively. The vectorization is based on the occurrence frequency of each aspect-based sentiment (within a specific period), which is analogous to the traditional bag-of-words method. We then compute the similarity of their profiles obtained from ASEP (*emb profile*) for comparison. In this case, V_A and V_B are the 128-dimension embeddings corresponding to employee A and B, respectively.

Additionally, to examine whether our profiles can capture any implicit social network information, we use t-SNE (Maaten & Hinton, 2008) to visualize both the *freq* and *emb profiles* in 2D space while tracking the position of the transformed specialist and two of her associates that she frequently communicates with.

Recall that ASEP identifies potential insiders based on their anomaly ranking. Hence, we examine how the anomaly ranking of the transformed specialist's profile has been affected by the augmentation in *Enron+*. More specifically, we perform two sets of experiments for both *freq* and *emb profiles*, one with the original *Enron* corpus and another with *Enron+*.

4.3. Results and discussion

4.3.1. ABSA results

The ABSA results are presented in Table 1. The results show that our ABSA model outperforms all the baseline models[4]. Furthermore, we observed that the results obtained here are similar to the tweets (social media channel) results reported in state-of-the-art ABSA approaches which are typically lower compared to the results for reviews (Liu, Zhang, Zeng, Huang, & Wu, 2018). This result is understandable, as sentences found in communication channels such as tweets or emails are more general and noisy, unlike reviews where the writer typical aim to express their views/sentiments towards certain aspects. The results also prove that incorporating restaurant and laptop reviews helps improve the accuracy of our model to some extent. Despite not having the perfect ABSA results, we demonstrate in the following that ASEP is still capable of capturing the emotional change in the induced scenario. When better ABSA technique is available we can replace it with our current model to improve the overall reliability of our framework. We also plan to explore and include other forms of emotion analysis into our framework as part of our future work.

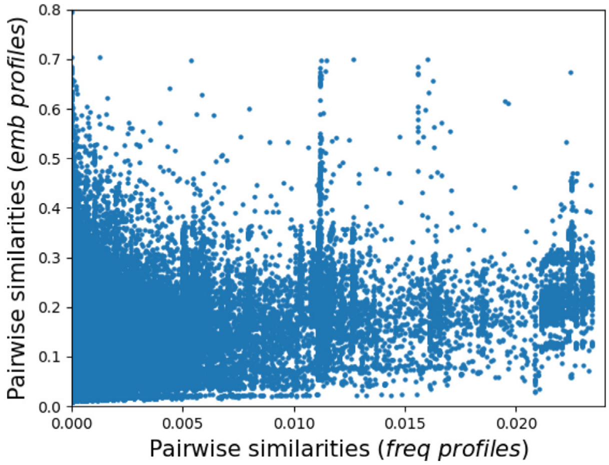


Fig. 6. Scatter plot to identify correlation between the similarities in aspect-based sentiment and our profiles.

4.3.2. Profile assessment

For the profile assessment, we focus on the January 2002 profiles, since it is the latest time interval where we have relatively more complete data available in the *Enron* corpus as by the first week of February 2002, most of the communications have ceased.

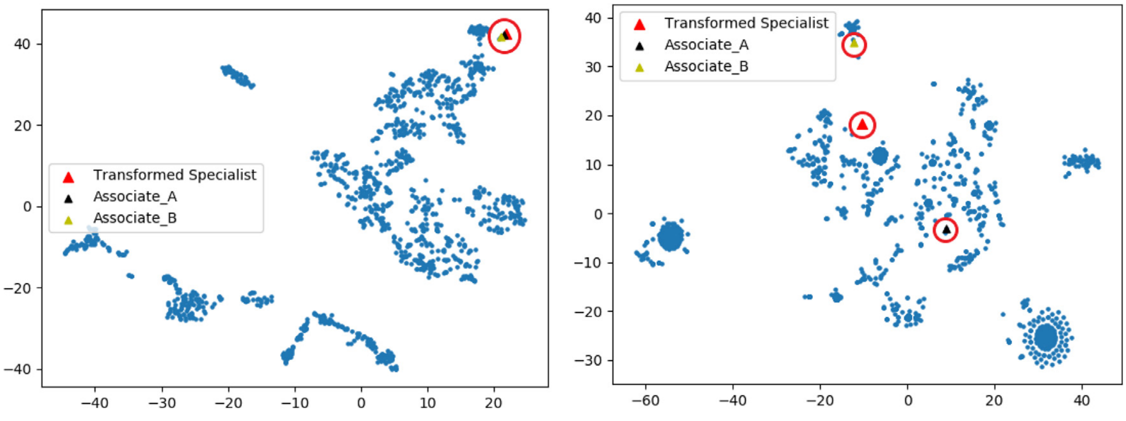
Fig. 6 shows the relation between the two sets of pairwise similarities obtained from *freq* and *emb profiles*. The occasional ”walls” (e.g., at 0.011, 0.016 and 0.023 of the x-axis) observed, occurs when multiple pairwise similarities in terms of aspect-based sentiment within a small range of values have a wide range of possible pairwise similarities values in terms of their profiles. However, from the figure, we can observe a general trend that as the pairwise similarities obtained from the *freq* increase, the pairwise similarities obtained from the *emb profiles* also increase. The Pearson correlation(r) between them is $r = 0.259$. Both the figure and Pearson correlation, show that the two sets of similarities have a weak positive linear relation. That is when two employees have similar aspect-based sentiment, their profiles also tend to be similar. The reverse is also true, for two employees who have dissimilar aspect-based sentiments, their profiles tend to be dissimilar. The main reason for the weakness in correlation could be due to the ability of the *emb profiles* to also capture the implicit social network information, which we will investigate in the following.

Fig. 7 shows the t-SNE plots based on both *freq* and *emb profiles*, obtained from *Enron+* in January 2002. In both plots, the transformed specialist and the two associates she often communicates with, are highlighted with different color markers and encircled in red. From Fig. 7a, it is clear that the three of them are closely positioned in the plot and that they obviously belong to the same clique while separated from others. On the other hand, in Fig. 7b, the plot obtained from *freq profiles* shows the three associates from the same clique to be widely separated. This is likely because the *freq profiles* are unable to comprehend the implicit social network information, analogous to how the bag-of-words model fails to capture the context information unlike word embedding methods (Mikolov, Sutskever et al., 2013). In other words, the clusters in Fig. 7a are based on the aspect-based sentiments and network information, while the clusters in Fig. 7b are based solely on aspect-based sentiments. Therefore, we can see that the points in Figure are more scatter compared to Fig. 7a.

4.3.3. Potential insider

Table 2 shows the monthly anomaly rankings for the transformed specialist based on both *freq* and *emb profiles*, from September 2001 to February 2002 in both *Enron* and *Enron+*. From

[4] https://github.com/songyouwei/ABSA-PyTorch.



(a) t-SNE visualization for *emb profiles*

(b) t-SNE visualization for *freq profiles*

Fig. 7. t-SNE visualization of January 2002 profiles.

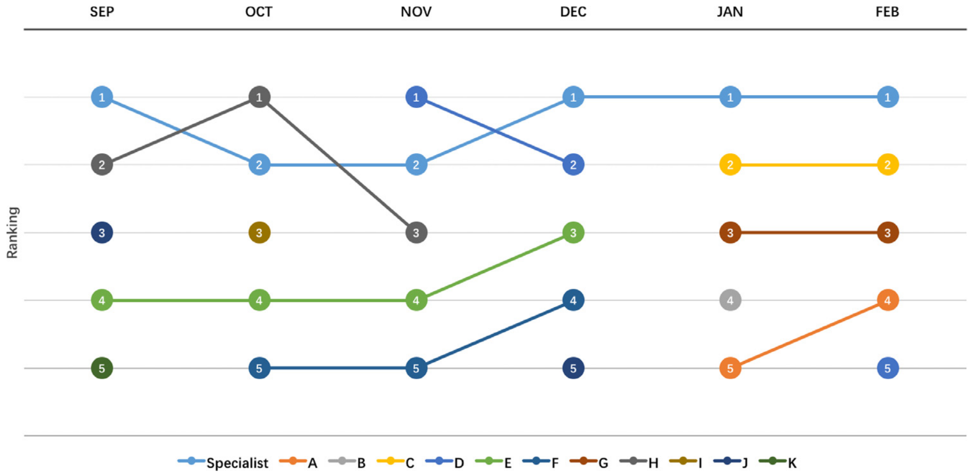


Fig. 8. Trend for top 5 ranking employees.

Table 2
Anomaly ranking for specialist.

Year		2001				2002	
Month		Sep	Oct	Nov	Dec	Jan	Feb
Enron	Number of Active Employees	2119	2423	1909	880	1212	920
	Ranking of Specialist (*freq*)	229	587	156	10	25	37
	Ranking of Specialist (*emb*)	518	802	738	173	154	18
Enron+	Number of Active Employees	2121	2433	1921	897	1240	935
	Ranking of Specialist (*freq*)	5	70	41	3	2	6
	Ranking of Specialist (*emb*)	1	2	2	1	1	1

the results, we can see that for both *freq* and *emb profiles* the rankings for *Enron+* are typically lower compared to the rankings for *Enron*. This suggests that our transformation of the specialist into an insider is reflected in the anomaly rankings. From the rankings for *Enron+*, we can see that the ranking given by *freq profiles* for October and November 2001 are low compared the other months. This could be due to the influence of noise resulting from the total number of active employees and amount of emails corresponding to each employee.

Furthermore, we analyze the top five employees in the anomaly ranking for these months. We chose to discuss the top five as we found that it is sufficient to observe the available trend while keeping the visualization simple. Similar observations were made when more top-ranking employees are analyzed. Excluding the augmented specialist, we have observed a total of eleven other employees appearing in the top five anomaly ranking over the six months. Due to privacy concerns the eleven employees are anonymized and represented by the capital letters, A to K.

Figs. 8 and 9 show the ranking trend and the occurrence frequency of these twelve employees (including the augmented specialist) in these months, respectively. In Fig. 8, when an employee appears for two or more consecutive months, a line is drawn to link the employee's previous ranking to the next ranking. A limitation of our current implementation is that is will always rank the employees according to their anomality, even when all the employees are actually normal. Therefore, appearing at the top of the anomaly ranking does not naturally means that the employee has high potential as an insider threat. As such, we recommend that the organization should spend more effort on observing the employees who are consistently and highly ranked in the anomaly ranking. From the two figures, we can observe that apart from the specialist who occupies the top two ranking consistently, most

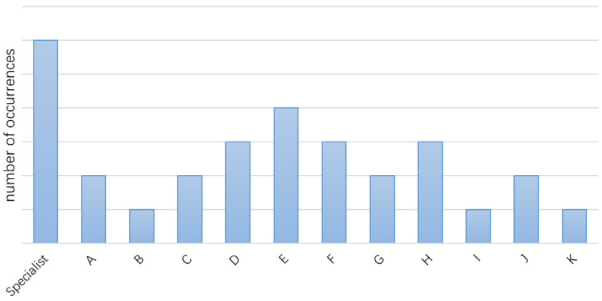


Fig. 9. Occurrence for top 5 ranking employees.

of the other eleven employees indeed only appear in the top five ranking occasionally and for a short period.

In sum, the results show that ASEP can effectively capture the anomalies in the employees' aspect-based sentiment and their implicit social network information. It encourages a new research direction to explore the value of ABSA in insider threat analysis, instead of the typical focus on negative sentiments or psychological characteristics.

5. Conclusions and future work

In this paper, we presented ASEP, a framework for identifying potential insider threats in an organization based on the employees' social communications (i.e., emails). As studies suggest that insiders typically went through a phase of emotional change before they commit the act, we propose to analyze the aspect-based sentiment of the employees and combine it with network information to build employee profiles for early detection. The ABSA module in ASEP performs sentiment classification on the aspect-based sentences extracted from emails. The employee profiling module leverages on the ABSA results and the network information and uses skipgram algorithm to learn the profile for each employee. Finally, the anomaly detection module gives each employee profile and anomaly score which are then ranked according to their anomality. The evaluation results show that by augmenting the emails of a particular *Enron* employee to show more negativity, her anomaly ranking rose to the top. Furthermore, the plot of the profiles in 2D space shows that the said employee and her two close associates are closely placed and separated from others. These show that the employee profiles obtained from ASEP can successfully encode the implicit social network information and more importantly their aspect-based sentiments

This work can be improved and extended in several ways. We discuss the potential future working direction in the following:

- Fully automated identification of relevant aspects. Currently, the aspect terms used for performing ABSA are chosen manually, as the automatically extracted aspects are mostly irrelevant. With the rapid advancement in the field of natural language processing, state-of-the-art language modeling technique can be adapted to automatically learn the relevant aspects for a given corpus.
- Enrich the profiles with additional emotional/ physiological information. In this work, the profiles are based on ABSA, however, social communications encompass much richer context such as the emotional (e.g., happy, sad, angry, etc.) or psychological information (e.g., personality traits, implicit motives, etc.). Such information can be leveraged to build more comprehensive profiles.
- Currently, the abnormality of an employee is computed with respect to the employee's difference in sentiments towards the aspects as compared to other employees. However, such sentiments could differ across department and position, which introduces noise into the system. Alternatively, we could also look into the abnormality of a particular employee compared to his/ her previous self to observe the change in his/ her emotional/ psychological state.
- The current framework works well on the dataset with a manually crafted insider scenario. However, future study in collaboration with psychologists and linguists to verify the correlation between psychological/emotional cues found in text and potential insider activities is paramount to validate the viability of the framework in the real-world

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eswa.2019.05.043.

Credit authorship contribution statement

Charlie Soh: Writing - review & editing, Software, Validation. **Sicheng Yu:** Software, Validation, Visualization, Investigation, Data curation. **Annamalai Narayanan:** Conceptualization, Methodology. **Santhiya Duraisamy:** Data curation, Writing - original draft. **Lihui Chen:** Writing - review & editing, Supervision.

References

Al Tabash, K., & Happa, J. (2018). Insider-threat detection using gaussian mixture models and sensitivity profiles. *Computers & Security, 77*, 838–859.

Alahmadi, B. A., Legg, P. A., & Nurse, J. R. (2015). Using internet activity profiling for insider-threat detection,.

Almehmadi, A., & El-Khatib, K. (2014). On the possibility of insider threat detection using physiological signal monitoring. In *Proceedings of the 7th international conference on security of information and networks* (p. 223). ACM.

Alotibi, G., Clarke, N., Li, F., & Furnell, S. (2016). User profiling from network traffic via novel application-level interactions. In *Internet technology and secured transactions (ICITST), 2016 11th international conference for* (pp. 279–285). IEEE.

Bagheri, A., Saraee, M., & de Jong, F. (2013). An unsupervised aspect detection model for sentiment analysis of reviews. In *International conference on application of natural language to information systems* (pp. 140–151). Springer.

Brown, C. R., Watkins, A., & Greitzer, F. L. (2013). Predicting insider threat risks through linguistic analysis of electronic communication. In *System sciences (HICSS), 2013 46th Hawaii international conference on* (pp. 1849–1858). IEEE.

Cambria, E., Fu, J., Bisio, F., & Poria, S. (2015). Affectivespace 2: Enabling affective intuition for concept-level sentiment analysis.

Camiña, J. B., Hernández-Gracidas, C., Monroy, R., & Trejo, L. (2014). The windows-users and-intruder simulations logs dataset (wuil): An experimental framework for masquerade detection mechanisms. *Expert Systems with Applications, 41*(3), 919–930.

Cappelli, D. M., Moore, A. P., & Trzeciak, R. F. (2012). *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)*. Addison-Wesley.

Chen, P., Sun, Z., Bing, L., & Yang, W. (2017). Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 452–461).

Chen, T., Xu, R., He, Y., & Wang, X. (2017). Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Systems with Applications, 72*, 221–230.

Chi, H., Scarllet, C., Prodanoff, Z. G., & Hubbard, D. (2016). Determining predisposition to insider threat activities by using text analysis. In *Future technologies conference (FTC)* (pp. 985–990). IEEE.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv:1409.1259.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555.

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 855–864). ACM.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2018). Insight into insiders: A survey of insider threat taxonomies, analysis, modeling, and countermeasures. arXiv:1805.01612.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980.

Klimt, B., & Yang, Y. (2004). The enron corpus: A new dataset for email classification research. In *European conference on machine learning* (pp. 217–226). Springer.

Kowalski, E., Cappelli, D., Conway, T., Willke, B., Keverline, S., & Moore, A. (2008). Insider threat study: Illicit cyber activity,.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196).

Legg, P. A., Buckley, O., Goldsmith, M., & Creese, S. (2014). Visual analytics of e-mail sociolinguistics for user behavioural analysis.. *Journal of Internet Services and Information Security, 4*(4), 1–13.

Leu, F.-Y., Tsai, K.-L., Hsiao, Y.-T., & Yang, C.-T. (2017). An internal intrusion detection and protection system by using data mining and forensic techniques. *IEEE Systems Journal, 11*(2), 427–438.

Liu, e., De Vel, O., Han, Q.-L., Zhang, J., & Xiang, Y. (2018). Detecting and preventing cyber insider threats: A survey. *IEEE Communications Surveys & Tutorials, 20*(2), 1397–1417.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth IEEE international conference on data mining* (pp. 413–422). IEEE.

Liu, Q., Zhang, H., Zeng, Y., Huang, Z., & Wu, Z. (2018). Content attention model for aspect based sentiment analysis. In *Proceedings of the 2018 world wide web conference on world wide web* (pp. 1023–1032). International World Wide Web Conferences Steering Committee.

Liu, Y., Corbett, C., Chiang, K., Archibald, R., Mukherjee, B., & Ghosal, D. (2009). Sidd: A framework for detecting sensitive data exfiltration by an insider attack. In *2009 42nd Hawaii international conference on system sciences* (pp. 1–10). IEEE.

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. arXiv:1709.00893.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research, 9*(November), 2579–2605.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Mohammad, S. M., & Turney, P. D. (2013). *Nrc emotion lexicon*. National Research Council, Canada.

Narayanan, A., Chandramohan, M., Chen, L., Liu, Y., & Saminathan, S. (2016). subgraph2vec: Learning distributed representations of rooted sub-graphs from large graphs. arXiv:1606.08928.

Pagliari, R., Ghosh, A., Gottlieb, Y. M., Chadha, R., Vashist, A., & Hadynski, G. (2015). Insider attack detection using weak indicators over network flow data. In *Military communications conference, MILCOM 2015-2015 IEEE* (pp. 1–6). IEEE.

Palus, S., Bródka, P., & Kazienko, P. (2010). How to analyze company using social network? In *World summit on knowledge society* (pp. 159–164). Springer.

Park, W., You, Y., & Lee, K. (2018). Detecting potential insider threat: Analyzing insiders' sentiment exposed in social media. *Security and Communication Networks, 2018*.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 701–710). ACM.

Ponemon, L. (2011). annual study: Cost of a data breach. ponemon institute.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)* (pp. 27–35).

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics, 37*(1), 9–27.

Sibai, F. M., & Menascé, D. A. (2011). Defeating the insider threat via autonomic network capabilities. In *Communication systems and networks (COMSNETS), 2011 third international conference on* (pp. 1–10). IEEE.

Song, Y., Salem, M. B., Hershkop, S., & Stolfo, S. J. (2013). System level user behavior biometrics using fisher features and gaussian mixture models. In *Security and privacy workshops (SPW), 2013 ieee* (pp. 52–59). IEEE.

Tang, D., Qin, B., Feng, X., & Liu, T. (2016). Effective lstms for target-dependent sentiment classification. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 3298–3307).

Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. arXiv:1605.08900.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Wu, Y., Zhang, Q., Huang, X., & Wu, L. (2009). Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 conference on empirical methods in natural language processing: Volume 3-volume 3* (pp. 1533–1541). Association for Computational Linguistics.

Zhu, R., Zeng, D., & Kosorok, M. R. (2015). Reinforcement learning trees. *Journal of the American Statistical Association, 110*(512), 1770–1784.