SUMMARY With the successful development and rapid advancement of social networking technology, people tend to exchange and share information via online social networks, such as Facebook and LINE. Massive amounts of information are aggregated promptly and circulated quickly among people. However, with the enormous volume of human-interactions, various types of swindles via online social networks have been launched in recent years. Effectively detecting fraudulent activities on social networks has taken on increased importance, and is a topic of ongoing interest. In this paper, we develop a fraud analysis and detection system based on real-time messaging communications, which constitute one of the most common human-interacted services of online social networks. An integrated platform consisting of various text-mining techniques, such as natural language processing, matrix processing and content analysis via a latent semantic model, is proposed. In the system implementation, we first collect a series of fraud events, all of which happened in Taiwan, to construct analysis modules for detecting such fraud events. An Android-based application is then built for alert notification when dubious logs and fraud events happen.

key words: Facebook, fraud analysis, latent semantic analysis, natural language processing, social networks

1. Introduction

With the universality of smartphones and the rapid development of mobile communication technology, a variety of social applications, such as Facebook, Line, Twitter, WeChat and WhatsApp, have been designed for smartphones in recent years. All of these social applications are able to significantly reduce the cost of communication among people, and make daily life more convenient than before. However, while taking advantage of the convenience afforded by social applications in life, many potential risks may arise. Generally speaking, two kinds of application risks exist with smartphones, i.e. mobile application risks and human behavior risks. In the first categorization, for example, many mobile applications ask users to agree to grant some high-risk permissions, such as “INTERNET” and “READ_CONTACTS”, and then perform these permitted privileges in the background without users’ scrutiny. A possible consequence of this is sensitive personal information being revealed via the Internet. On the other hand, from the perspective of human behavior risk, fraudsters may use tricks on social media to communicate with victims for specific fraudulent purposes. For instance, fraudsters may pretend to a friend of a victim and chat with him/her via Facebook or Line for the purpose of tricking the victim into offering fraudsters some virtual currency, i.e. online game points, or even to steal his/her credit card’s serial number and card security code. Huge losses may be incurred by victims who fall for such ruses. In 2015, according to the Taiwan government’s official report provided by the Criminal Investigation Bureau[15], more than 8 thousand fraud events were reported and around 670 million NT dollars was lost in Taiwan. It is thus vitally important to heighten awareness of personal information protection, including the required know-how to prevent falling victim to fraud. In the literature, adaptive fraud detection [6], signature-based systems [2] and host-based intrusion detection systems [3] have all been proposed to detect and prevent suspicious events and fraudulent behaviors via network monitoring and packet flow analysis. Unfortunately, these methods may not provide accurate discernment when it comes to protecting against fraud events launched by a legitimate but malicious fraudster. For example, a fraudster may register a legitimate account and then exploit social engineering tricks during real-time messaging communications operated on smartphones to cheat victims. In this case, network-level monitoring is not enough to detect a fraud event, and a content-based analysis is required.

Frauds perpetrated on social networks usually happen on smartphones and involve a series of malicious steps of social engineering being launched. Well-known examples involve bogus insurance claims, tax return claims, credit card transactions and online purchases. To effectively detect and predict fraud events on social networks, data analysis techniques are vitally important. In general, analysis techniques for fraud detection can be classified into two major classes: statistical-based techniques and machine learning-based techniques. Examples of statistical data analysis techniques are data pre-processing, analysis on statistical parameters, and prediction and detection via operation patterns against activities and user profiles, all of which are intended to estimate risks and predict fraud. On the other
hand, machine learning-based analysis techniques are ones such as data mining, expert systems, pattern recognition and neural networks. As it stands, data analysis techniques have been widely used to detect fraud events on social networks [5], [7]–[13]. However, it is difficult to effectively and efficiently detect and prevent fraud on social networks since fraud is a complicated and adaptive crime process which involves various areas of knowledge, such as biology, financial and business operation. In addition, to detect fraud events on social networks, individuals (or organizations) usually need to monitor and analyze transactions against critical system parameters and data analysis patterns. This requires complex and time-consuming investigations involving the analysis of enormous transaction logs.

In this paper, we would like to introduce a novel analysis method for fraud identification on social networks. In contrast to existing methods, we present a fraud detection method in which the fraudulent features are extracted from a chat log via text mining analysis. We utilize natural language processing (NLP) to filter out the meaningless data and further adopt latent semantic analysis (LSA) to eliminate the noise signal from the analyzed data set. The fraudulent feature is then extracted and used to identify the fraud event. In brief, the main contributions of this study are as follows: (1) an integrated platform with text-mining techniques is developed to prevent the deceptive fraud events in the real world; (2) the proposed method is efficient for real-time fraud detection on social networks and possesses high accuracy of fraud detection as well. The rest of the paper is organized as follows. Section 2 introduces the state of the art of fraud detection on social networks. In Sect. 3, we present the proposed fraud detection system with detailed explanations of each data analysis module. Then, we demonstrate the results of the tested scenarios and the findings gathered during the evaluation in Sect. 4. Finally, we give a concluding remark in Sect. 5.

2. Related Works

In this section, we introduce the state of the art of fraud detection. In 2009, Yu and Wang [13] proposed a fraud detection model via an outlier judgement process. A distance sum between the target record and other object records is computed by the Euclidean distance against the infrequency and unconventionality of the fraud observed in a historic credit card transaction dataset. According to a higher degree of unexpectedness and a longer time interval, suspicious records can be identified. The authors claimed that outlier mining is superior to clustering-based anomaly detection for credit card fraud analysis and identification. Zhu et al. [11] observed that the buyer-feedback mechanism is not sufficiently effective when it comes to seller reputation verification, and accordingly provides weak fraudulent behavior prevention. A fraud analysis and detection scheme was constructed via the relationships in social networks during transactions. In more detailed terms, the relationships between users in an online auction system were investigated to verify whether credit speculation happened or not. The authors examined the feasibility of the proposed scheme via the evaluation of real data from a famous electronic commerce platform in China, i.e. Taobao. Next, Ying et al. [12] demonstrated a spectrum based analysis method to detect the launching of a malicious attack, called random link attacks. In the attack scenario, a malicious user may create multiple fake identities and then launch malicious tricks targeting regular members of the network via interactions disguised as legitimate user behaviors, while using those fake identities. The authors then presented an analysis scheme which exploits the spectral space of the underlying network topology to identify frauds or attacks. The spectral characteristics of a potential attacker are determined by the regular users.

In 2012, Jamshidi and Hashemi [5] proposed a data enrichment mechanism embedded within social network analysis to detect fraud scenarios. The presented method concentrates on the efficiency of update procedures when new users (or new transactions) emerge. The probability of a fraud is calculated based on the relations between the tested event and known frauds. Next, Sylla et al. [8] combined several methods and tools to solve the problems of linking information spread across different heterogeneous data repositories for detecting fraud and other crimes targeting banking and consumption on social networks. Later, Nandhini et al. [7] proposed a methodology to identify fraudulent activities and suspicious profiles on social networks. Their proposed method can help online users to safely communicate with each other by preventing particular malicious activities, such as profile hacking and phishing attacks. Then, Agrawal et al. [1] combined several techniques, such as a genetic algorithm, a behavior based technique and a hidden Markov model, to detect fraud events during the processes of registration, login, banking and proceeding with shopping. After that, Wu et al. [10] introduced a continuous authentication scheme to detect in-situ identity fraud, in which an attacker operates as a legal user with the victim’s account information, and even, in some instances, with the victim’s own devices. The presented method analyzes and detects the in-situ fraud via patterns constructed from users’ browsing behaviors, such as habits of photo viewing and page switching. The authors demonstrated that their method can provide greater than 80% detection accuracy within 2 minutes, and 90% after 7 minutes of observation time on the Facebook platform. In the same year, Vlasselaer et al. [9] introduced a fraud detection approach, called active fraud investigation and detection, which uses active inference to effectively detect fraud in time-varying social networks. In active inference on social networks, a set of unlabeled events is considered and marked as fraud by inspectors. All of the marker labels will then be used in the inference process by the trained classifiers. In the proposed method, fraudulent and non-fraudulent events are first classified and then the analysis is performed via intrinsic features and neighborhood features extracted from a time-varying network.
3. The Proposed Fraud Detection System

The goal of this study is to develop a fraud analysis system which can efficiently and effectively detect fraud events during real-time messaging communications on social networks. In this section, we describe the details of the proposed system consisting of five modules, i.e., data collection, NLP, matrix processing, semantic models and similarity evaluation. The system architecture is shown in Fig. 1. First, we collect the existing fraud events in Taiwan as the fraud templates for the purpose of data analysis. Note that all of these data are real fraud scripts which happened in Taiwan. Next, we perform the NLP technique on the testing data, in which word segmentation and the elimination of stop words and special symbols are executed. Then, we put the processed data into the matrix processing and transform it into a VSM (Vector Space Model) and TF-IDF (Term Frequency-Inverse Document Frequency). Finally, based on the LSA (Latent Semantic Analysis) model and similarity evaluation, the results of fraud detection are presented.

3.1 Data Collection

In the data collection phase, we first collect the relevant news about fraud events via search engines, i.e., Google and Bing. In addition, we obtain data on real cases of fraud event through the Taiwan government’s anti-fraud center website [15]. The two kinds of source data are integrated to make up the fraud templates during our implementation. Next, in order to transform the original testing data from the search engine into predefined tested-format data, we perform a series of NLP techniques.

3.2 Natural Language Processing (NLP)

This section is divided into three steps, i.e., Chinese word segmentation and the elimination of stop words and special symbols. First, the system performs a Chinese word segmentation operation via CKIP service [14]. CKIP word segmentation can be exploited to retrieve information and, in particular, to retrieve critical keywords appearing in the testing data. Second, in general, stop words usually refer to the most common words in a language such as particles, adverbs and conjunctions. It is thus suggested to remove stop words from the testing data to prevent deviation in the analysis results. Hence, we adopt the stop word lists provided by Academia Sinica of Taiwan to remove stop words from the testing data. Third, since Chinese sentences are always composed with punctuation marks, such as commas, periods, quotation marks, and brackets, we establish a list of special symbols for removing punctuation marks and special symbols from the testing data.

3.3 Matrix Processing

Before performing the LSA technique, we have to transform the processed testing data, which is the result of the NLP phase, as shown in Sect. 3.2, into VSM matrix form. We present the fraud templates adopted in this study in Table 1 and the VSM matrix of the fraud templates and the pre-defined fraud keywords in Table 2. Based on the VSM matrix, a TF-IDF matrix is then generated. The TF-IDF frequency increases proportionally with the number of times a word appears in the testing data. However, the word’s importance is offset by its frequency in the corpus, so it will be adjusted.

Term Frequency (TF): The importance of term $t_i$ in the testing data can be presented as Eq. (1).

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$  \hspace{1cm} (1)

Note that $n_{i,j}$ is the number of times the term $t_i$ appears in the testing data $d_j$, while $\sum_k n_{k,j}$ is the total number of times the term appears $t_i$ in the testing data $d_j$.

Inverse Document Frequency (IDF): The IDF is able to measure the amount of information provided by the word. It is the logarithmic scale of the testing data containing the keyword, which is obtained by dividing the total number of documents by the number of documents containing the keyword.

<table>
<thead>
<tr>
<th>Lex</th>
<th>Content</th>
</tr>
</thead>
</table>
| 41  | 提問意見
| 42  | 香港金融管理局 (HKMA) 之基本資料
| 43  | 財政事務司司員 資訊及知識產品
| 44  | 中外貿易發展
| 45  | 金融管理局 (HKMA) 之基本資料
| 46  | 財政事務司司員 資訊及知識產品
| 47  | 中外貿易發展
| 48  | 金融管理局 (HKMA) 之基本資料
| 49  | 財政事務司司員 資訊及知識產品
| 50  | 中外貿易發展
| 51  | 金融管理局 (HKMA) 之基本資料

Fig. 1 The proposed fraud detection system
Table 2  The VSM matrix of the fraud templates and the pre-defined fraud keywords

<table>
<thead>
<tr>
<th>term</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
<th>d9</th>
</tr>
</thead>
<tbody>
<tr>
<td>浪钱</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>利</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MY CARD</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>電話</td>
<td>8</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>手機</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>送信</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>促销</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>刺激</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>驚</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>预防</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>疑问</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>问</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>半</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>惠</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>票面</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>票数</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>談及</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>討</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>通行</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>精</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>而</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>鉅</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>付</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>並及</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>利潤</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>竹</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>代</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>销</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>财</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3  Matrix $A$

<table>
<thead>
<tr>
<th>$A_{ij}$</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
</table>

3.4 Latent Semantic Analysis (LSA)

LSA is a model based on mathematical statistics combining singular value decomposition (SVD) with dimensionality reduction. LSA can not only depict the information about the testing data, but can also derive the relationships between latent semantics and information. In this section, given a real matrix $A_{m \times n}$ and supposing $m \geq n$, $\text{rank}(A)=r$, $r$ means the quantity of singular values, and represents the rank size of the matrix as well. Operating singular value decomposition for $A_{m \times n}$ can be expressed as a formula, as shown in Eq. (4). Table 4 shows the result matrix $\bar{A}$ of the LSA operation.

$$A = U \Sigma V^T$$

$U$: Term vector matrix.
$\Sigma$: Retaining singular value matrix.
$V^T$: Vector matrix.

3.5 Cosine Similarity

In this section, the relevance of the testing data and the fraud templates is evaluated via cosine similarity. Cosine similarity is always adopted in positive space, where the outcome is bounded between $[0, 1]$. In the area of information retrieval and text mining, each term is notionally assigned a different dimension, and a document is characterized by a vector where the value of each dimension corresponds to the number of times that term appears in the document. Cosine similarity is able to provide a useful measure of how similar two documents are likely to be, in terms of their subject matter. The cosine of two vectors can be derived by using the Euclidean dot product formula as presented in Eq. (5).

$$a \cdot b = \|a\|\|b\|\cos \theta$$

In that case, given two vectors of attributes, $A$ and $B$, the cosine similarity, $\cos(\theta)$, is represented as in Eq. (6).

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|a\|\|b\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

The resulting similarity ranges from $-1$ meaning exactly opposite, to $1$ meaning exactly the same, with $0$ indicating orthogonality (decorrelation), and in-between values.
indicating intermediate similarity or dissimilarity.

4. Implementation: Scenario Testing

In this section, we develop a demo-system to present the practicability of the proposed fraud detection method. Figure 2 shows the implementation of the proposed fraud detection system. An Android-based application is implemented on a NEXUS 9 tablet to support chat log retrieval and alert notification, while the backend server is operated on a desktop computer and responsible for data pre-processing, matrix processing and data analysis. The normal procedure of our proposed system is as follows. First, the message interception module in the Android-based application is used for retrieving the real-time chat logs from Facebook Messenger, and the chat logs are then forwarded to the backend server. Next, the server performs NLP procedure on the chat logs via the message processing module and converts the results into a TFIDF matrix via the matrix processing module. After that, LSA is operated at the semantic analysis module to eliminate the noise signal during the data analysis phase. Finally, the decision module computes the relevance between the testing data and pre-defined fraud scripts, and returns the fraud detection result to the Android-based application. The warning notification module then sends a notification to inform the user whether a fraud event is happening (or has been correctly identified) or not.

In our experiment, we investigate fraud information from a search engine and anti-fraud website to simulate two kinds of real fraud tricks: one group (including d1 to d6) involves a game-point based fraud event, and the other group (including d7 to d9) involves a micro-payment based fraud event. Note that in our experiment d1-d9 are adopted as fraud templates which really happened in Taiwan. Regarding the testing data, we simulate three kinds of groups: (1) one group consists of fraud events, i.e. t1 to t5; (2) one group consists of non-fraud events with fraud keywords, i.e. t6 to t8; and (3) the others are non-fraud events without any fraud keywords, i.e. t9 and t10.

Figures 3 and 4 present the snapshot of our developed Android-based application on a NEXUS 9 tablet. In Fig. 3, a smartphone user opens our fraud detection application and clicks the icon to monitor (and retrieve) chat logs from Facebook Messenger, as shown in Fig. 4. Next, the applications sends retrieved real-time messaging (i.e. a chat log) to the backend sever, which then executes the NLP and LSA techniques on the chat log for fraud analysis and detection. Then, the application receives the analysis result from the backend sever and sends a notification to the user (such as that shown in Fig. 5).

In our experiment, we set up two processes to judge whether an instance of testing data is fraud or not. First, we calculate the cosine similarity between the TF-IDF matrices derived from the testing data and the pre-defined fraud template, respectively. The TF-IDF similarity ranges from zero (irrelevant) to one (totally relevant). Next, we further calculate the cosine similarity between LSA-performed matrices, which is based on the TF-IDF matrices in the first step. If the LSA similarity is greater than the TF-IDF similarity and a pre-defined threshold, it means that the testing data is probably a fraud event. We present the following two scenarios to test the performance of the proposed system.

Scenario 1 (fraud-event): In scenario 1, we input a real fraud event as the testing data. Given an input data t1 as shown in Fig. 5, we can derive the TF-IDF similarity between t1 and the fraud templates (i.e. d1-d6 and d7-d9) as shown in Table 4. In addition, the LSA similarity between t1 and d2, d5 and d6, respectively, is greater than the TF-IDF similarity and the pre-defined threshold. This means that the input data t1 is probably a fraud event.

Scenario 2 (non-fraud event with fraud keyword): In scenario 2, we input a testing data which is a non-fraud event
with a fraud keyword. Given an input data $t_2$ as shown in Fig. 5, we can derive the TF-IDF similarity between $t_2$ and the fraud template, i.e. $d_1$-$d_9$, as shown in Table 5. Nevertheless, the LSA similarity of $t_2$ between the fraud templates (i.e. $d_1$-$d_6$) is less than the TF-IDF similarity and the predefined threshold. This means that the input data $t_2$ is probably not a fraud event and an alert notification is shown as appears at the top of Fig. 4.

Finally, we summarize the performance results in Table 6, in which ten testing data are adopted. We can see that the results of testing data ($t_1$-$t_5$, $t_9$, $t_{10}$) are judged accurately, as expected. Nevertheless, the result of testing data ($t_6$, $t_7$) is not as expected. We observe that there are many negative values between terms and data in the matrix produced by LSA. The negative values certainly affect the result. Thus, we need to consider both the TF-IDF similarity and the LSA similarity to help us judge whether the testing data is a fraud event or not.

5. Conclusion

Nowadays, people spend much of their time on social networks in which real-time messaging applications are used for communication. In this study, we focus on fraud analysis and detection during real-time messaging communications. The natural language processing and semantic analysis model is integrated and utilized to increase the effectiveness of fraud detection. We then build a demo-system to present the practicability of the proposed idea. To the best of our knowledge, we are among the pioneers when it comes to investigating fraud analysis and detection via user conversations. In the future, different semantic models, such as probabilistic latent semantic analysis (PLSA) and latent dirichlet allocation (LDA), may be adopted to enhance the effectiveness of fraud detection.

Acknowledgments

This work was supported in part by the Academia Sinica, in part by the Taiwan Information Security Center, and in part by the Ministry of Science and Technology, Taiwan under Grant MOST 105-2221-E-259-014-MY3, Grant MOST 105-2221-E-011-070-MY3, Grant MOST 105-2923-E-182-001-MY3, Grant MOST 105-2218-E-001-001 and MOST 106-3114-E-011-003.

References


Ping-Hsien Lin received the B.S. and M.S. degrees in information management from National Taiwan University of Science and Technology, Taipei, Taiwan, in 2013 and 2015, respectively. He is currently a software engineer and works on Wistron Information Technology & Services, Taipei, Taiwan. His master’s thesis is “A fraud detection system for real-time messaging communication on Android Facebook messenger”. His research interests include android mobile application security and website development of E-commerce.