Human-centric Computing and Information Sciences

January 2021 | Volume 11

www.hcisjournal.com

KIPS Korea Information Processing Society

KIPS CSWRG Korea Information Processing Society Computer Software Research Group

HCIS | Human-centric Computing and Information Sciences

RESEARCH(Research Manuscript) **Open Access**

Human-centric Computing and Information Sciences (2021) 11:02
DOI: https://doi.org/10.22967/HCIS.2021.11.002
Manuscriptrevised September 26, 2020; accepted December 23, 2020; published January 29, 2021

Voice Recognition and Document Classification-Based Data Analysis for Voice Phishing Detection

Jeong-Wook Kim[1], Gi-Wan Hong[1], and Hangbae Chang[2,*]

Abstract

Phishing crime has become a serious issueworldwide. Damagescaused by phishing have been increasing continuously ever since the first phishing attacks occurred. Voice phishing, in which criminals impersonate financial institutions over the telephonein order to damage consumers, account for the majority of such attacks.This study aimed to convert phishing sound source files to text files through voice recognition and to classify and evaluate whether such texts can be judged as voice phishing. From the proposed methodology, it was confirmed that the Doc2Vec embedding method and the similarity determination method performed better than the methods used in previous studies. Through this, the proposed methodology confirmed that voice phishing can be judged by document data that are textualized by voice recognition for voice phishing sound sources.

Keywords

Phone Scam, Voice Phishing, Natural Language Processing, Voice Recognition Document, Voice Detection Classification, AI, Machine Learning

1. Introduction

Non-face-to-face transactions have the advantage of being easy to conduct without regard to the place of business or the opening hours of financial institutions, whereas e-financial scams target financial consumers by overly simplifying the authentication process and the transaction process for using financial services.There are also various types of electronic financial fraud including phishing, pharming, and memory hacking. Voice phishing, in which criminals seek to damage consumers by impersonating financial institutions over the telephone, is the most important of these types.

***Corresponding Author:**Hangbae Chang(hbchang@cau.ac.kr)
[1]Dept. of Security Convergence, Graduate School, Chung-Ang University, Seoul, Korea
[2]Dept. of Industrial Security, Chung-Ang University,Seoul, Korea

The damages caused by voice phishing are increasing every year, reaching an all-time high ofUS$550 million in 2019 in South Korea. Because it is difficult to assess the full scale of the damage, even in the case of middle-aged and vulnerable people who are vulnerable to electronic financial fraud, it is highlylikely that the amount of damageswill be much higher. Considering that the damage caused by voice phishing is not only economic but also social damage, such as the overall health and welfare of the victims and the increasing rate of suicide attempts, it is vital to come up with countermeasures to this problem (Table 1).

As a countermeasure, the government has mandated the establishment of electronic financial fraud prevention services for all financial institutions since 2013, and announced joint measures to prevent voice phishing, and prepared countermeasures for omnidirectional voice phishing, in 2018. These efforts appear to reduce the damages caused by voice phishing in the early stages of their implementation, but it is difficult to determinewhether criminals who have identified the limitations of preventive services have changed their methods, and hence to eliminate completely the damages caused by voice phishing.

Table 1. Voice phishing damage status

Crime type	2015	2016	2017	2018	2019[a)]
Damage cost(US$ million)	202	159	201	367	332
Number of cases	57,695	45,921	50,013	70,218	38,068
Number of victims	32,764	27,487	30,919	48,743	26,086

Source from the Korea Financial Supervisory Service.
[a)]Statistics for the first half of 2019.

Prior studies related to voice phishing have largely consisted of an actual status analysis or policy studies from a legal and institutional point of view, and thus have limitations in establishing a system for relievingvictims of voice phishing. There are some technical case studies on methods of preventing voice phishing, but most studies have attempted to classify the phonetic characteristics of voice phishing criminals or to analyze the structural characteristics of voice phishing scripts, and there has been insufficient research and development on ways of preventing voice phishing.

Recently, attempts have been made to prevent voice phishing based on data analysis and artificial intelligence (AI), centered on financial companies. The real-time call analysis system consists in analyzing the user's call contents in real time by using artificial intelligence and calculating the probability that a given call is a voice phishing call. The core features of this process are that voice is converted to text in real time; natural language processing analyzes converted text based on morphemes; and text classification technology judges voice phishing based on expressed text. Although this AI-based voice phishing classification attempt shows utility from a practical point of view, it lacks research on accuracy analysis or the constraints of the algorithms used.

For the purposes of this study we performed a voice phishing call classification experiment, checked various technical constraints on speech recognition, natural language processing, and text classification used for classification, and compared the accuracy of a number of recently researched algorithms to prevent voice phishing. To further improve the accuracy of the system, we intend to study various technical methods.

This study is limited to experiments on models for verifying voice phishing crimes using Korean telephones and the classification of voice phishing using the Korean language among cases of electronic financial fraud in Korea. Other countries, including the United States and Japan, have continued to experience e-financial fraud, which has become a serious social problem; however, due to differences in language and the infrastructure of the Internet, which act as adifficult factor in voice phishing sound analysis, this aspect will be addressed infuture research [1]. As such, the scope of this

paper was limited to cases of electronic financial fraud involving Korean financial companies since 2006. In addition, the scope of the experiments conducted in this study was restricted to Korean phone records because the methods of analyzing Korean language, which has the characteristics of an agglutinative language, different considerablyfrom those used to analyze English and Chinese[2].

To date, related works aimed at preventing electronic financial fraud encompass voice phishing prevention and damage relief, the improvement or linkage of the abnormal symptom detection systems of financial companies to prevent the transfer of funds, and the addition or improvement of certified media in the process of security certification [3–5]. Most studies are focused on preventing social engineering approaches to voice phishing. Technically speaking, research is now needed to prevent e-financial fraud; and this study is different in that it attempts to compare the performance of different AI-based methods of detecting e-financial accidents [6]. This paper consists of experiments involving the detection and prediction of voice phishing, which accounts for the largest proportion of cases of electronic financial fraud, and performs data analysis using actual voice phishing sound source data to compensate for the shortcomings of previous studies, based on theoretical assumptions. The paper proposes a method of improving detection performance by reproducing the AI-based call analysis currently under development and by comparing the accuracy of its speech recognition, natural language processing and document embedding.

The rest of this paper is organized as follows: Section 2 reviews the related works;Section 3 proposes a model for voice phishing detection and describes the experiment; Section 4 presents the results of the experimentand the analysis of the performance of the proposed research model; and Section 5 presents the conclusion andthe direction of future research.

2. Related Works

2.1 Types of Voice Phishing Crimes

The initial method of voice phishing was a simple scam for stealing financial information by impersonating public institutions such as the National Tax Service under the pretext of tax refunds (automated teller machines), but other scamssuch as fake college tuition refunds and prize winners, emerged as the method evolved.In order to deceive a victim into believing a call pretending to be froma financial institution or government agency, the fraud method and speech method further evolvedinto the sending of a large number of mobile phone text messages using personal information obtained through illegal channels in advance, under the pretext of the illegal use of the victim's credit cards by third parties[1].

Crime by voice phishing can be divided into attacks targeting unspecified people and attacks targeting specific people using the personal information collected from the victims. As regards voice phishing targeting unspecified people, it can be classified into the "government agency impersonation type" and the "loan type" according to the methods of fraud used, while voice phishing targeting specific people can be classified as "threat", "duty overrun", and "personal impersonation" [7] (Table 2).

Table 2. Voice phishing type classification

Attack range	Fraud type	Impersonation agency	Scam story
Unspecified	Impersonation of governmentagency	Police, prosecution,court	Crime investigation implications, personal information leakage, etc.
	Loan fraud	Bank, credit card company, Financial Supervisory Service	Low interest rate solicitation, sovereignty loan conversion

Target	Blackmail	Violent organization	Mail, courier service, card return, etc.
	Obligatory sentence	Alumni Association, University	Request for membership fees, university admission, payment of tuition fees, etc.
	Impersonation	Family, friends, colleagues	Request for power supply, settlement amount, etc.

In the case of “government agency impersonation type”, it refers to a criminal method in which a victim transfers deposits and cash to a criminal’s fake bank account,after being deceived into believing that he/she is a suspect in a crime by a criminalpretending to be an employee of the prosecution, police, or the Financial Supervisory Service[1]. The first step is to impersonate a public institution such as the prosecutors’ office or a court, creating anxiety that the victim is involved in a crime and must appear in court within the deadline. Then, after asking about the possibility of leaking personal information, such as whether or not you have lost your ID, ask the victim for your personal information and account number to obtain the information. After obtaining the victim’s personal information and account numbers, people will be lured to automated banking devices for account protection measures by the Financial Supervisory Service. Finally, it is a fraudulent technique in which another criminal pretends to be an employee of the Financial Supervisory Service,sets a “safety code” on the account, and then manipulates an automated device in order to receive a transfer.

In the case of “loan type”, the victim is deceived that they can get a loan at a low interest rate or switch to another loan at a lower interest rate by creating your transaction performance and credit score. And then the criminals committed an extortion of loans or fees. In the case of damages, the loan is defrauded in the name of issuing a certificate of payment to the victim after receiving a high-interest loan. The criminal impersonates an employee of credit institution, deceives the victim that he must have a record of high-interest loans in order to obtain a low-interest loan with government-funded funds and gets the victim to take out loans from lenders[1].Subsequently, the criminal deceives the victim that the certificate of payment for loan repayment should be issued and sent to the credit institution. The criminal then makes the victim transfer the loan to the borrowed-name account and steals it.

Recently, new types of voice phishing which take advantage of the coronavirus disease 2019 (COVID-19)crisis have emerged. Voice phishing criminals defraud money by impersonating financial institutions and deceiving the victims into believing that low-interest, government-backed loans are possible. Examples include cases in which victims are tricked into stealing money by taking out a new loan to pay off an overdue loan; cases in which low-interest loans are taken out by demanding work costs for credit rating upgrades; and cases in which malicious applications are installed under the pretext of non-face-to-face loan process. The number of cases of intelligent voice phishing aimed at the vulnerabilities of financial institutions that could arise from non-face-to-face transactions is also on the rise [8].

In the case of “abduction and blackmailing”, there are cases in which criminals demand direct remittance afterclaiming to have kidnapped a member of the victim’s family. Generally, after obtaining the phone number and home number of the victim’s family via the Internet, such criminals keep callingand bothering the family via their cell phone[6]. After that, they call the home and threaten to kill the victim if the money isn’t deposited. At this time, the victimsare put under great pressure by the criminals, who create an atmosphere of fear by creatingthe soundsof children crying after being beaten.

As regards the “family or relative camouflage type”,criminals disguise themselves as relatives and deceive the victims intodepositing money into their bank accounts because they are in a hurry[1]. In this way, it is difficult to evaluate the damage due to late recognition, because the main victim is usually an elderly person. Even if one notices the damage, it is difficult to recover due to the delayed response. After calling the home, the criminal tells the victim that he has kidnapped a member of the victim’s family or friends. He threatens to murder the abductee if he/she does not deposit the money,

and forces the victim to transfer his/her savings.

As for the “type of aiming at jobseekers”, this is a method whereby criminals induce jobseekers to withdraw funds. As a part of the comprehensive voice phishing countermeasures established in 2018, the delayed withdrawal system was introduced in order to make it difficult to secure fake bank accounts. As a result, criminals then started to deceive jobseekers by opening bank accounts under the pretext of finding a job and devised fraudulent methods of inducing them to withdraw funds[1].

2.2 Analyzing Non-verbal Characteristics

An early study involving an analysis of voice phishing was conducted by analyzing the non-verbal characteristics of voice phishing voices to study the differences between voice phishing criminals and the general public [9]. In a study on a voice phishing detection algorithm using the minimum classification error technique, a detection algorithm based on the minimum classification error technique was proposed to extract non-verbal leakages (such as trembling voice, eye movements, hand movements, and subtle changes in facial expression) from sound sources [10].However, the paper had a limitation in that fewer than 10 voice phishing data were used to verify the model in the experiment, and only non-verbal leaks were analyzed without using the voice call history to determine voice phishing.

2.3 Machine Learning Techniques for Text Classification

2.3.1 Text classification

Text classification was performed by selecting the important features of each document using such methodologies as document frequency, information gain, mutual information, and χ^2 statistics as a statistically-based method of feature selection [11].Class-specific features were separately defined using thesupport vector machine (SVM), and documents were classified using the Bayesian classifier [12]. Some studies consisted in categorizing documents using the n-gram-based bag-of-words method as a feature selection method [13–15],while othersaimed to derive meaningful classification results through feature selection for natural language using the graph-of-words methodology, including research on a feature selection methodology for text categorization focused on naive Bayesian classifiers [16, 17].

2.3.2 Latent semantic analysis

Latent semantic analysis (LSA) refers to a methodology that reduces the number of dimensions through singular value decomposition for embedded text data and draws hidden meanings in the process [18]. With LSA, we can identify the structure of semantic relationships between words used for the statistical analysis of large groups of documents [19]. The general LSA model is based on singular value decomposition (SVD), which can reflect the basic structure of the various dependencies that exist in the original matrix. LSA extracts the latent meaning of the document by considering the meaning of the word and performs the similarity calculation of the document. Text classification research through latent semantic analysis is being carried out on an ongoing basis, and the Word2Vec-based latent semantic analysis model has been designed using K-means clustering as a topic modeling study to understand the trend of blockchain technology, and performance is compared using a general LSA model. Another study was conducted to prove its usefulness [20]

2.3.3 Word2Vec

Word2Vec can efficiently express large amounts of text data as vectors by using a simple artificial neural network model, requires little preprocessing, and processes various natural languages such as

sentiment analysis and machine translation because it has the advantage that the expressed word vector contains contextual information [21, 22]. The Word2Vec model helps determine the semantic distance between two words. It includes the continuous bag-of-word (C-BOW) model, which predicts the next word from context words as large as the surrounding window size, and the skip-gram model, which predicts surrounding context words from one word [23].

2.3.4 Doc2Vec

Doc2Vec is a technique for embedding a document through the Word2Vec technique. Doc2Vec extends Word2Vec from word level to the document level, so that each document has its own vector value in the same space as the word [24]. It represents similar documents represented as vectors close to each other. It calculates the probability of matching the next word in a given sequence of words. This model slides through the document and embeds the document in a way that predicts what the next word will be. It determines a variance representation for a single document by learning a neural network with information on the surrounding words of the target word. Doc2Vec includes two models: distributed memory (DM) and distributed word collection (DBOW) [25], of whichthe former predicts the probability of a word through context and document vectors, while the latter predicts the probability of an arbitrary set of words in documents with document vectors. In the actual document environment, since there are more text documents without labels and unstructured, multi-co-training was conducted, text classification was performed by using term frequency–inverse document frequency (TF-IDF), latent Dirichlet allocation (LDA) and Doc2Vec together to improve the classification accuracy of documents compared to the semi-supervised learning (SSL) approach [24].

2.4 KMP Algorithm-based Keyword Matching

The proposed method converts the phone voice into text in a phonecallthat is suspected of voice phishing and alerts the user when a key word used in voice phishing is heard in voice during a call. Thus, a situation in which voice phishing is suspected during a real-time call is generated as a warning to notify the user. New voice phishing patterns are being continuously updated to ensure the reliability of the patterns. In addition, the text conversion of the phone voice can be converted into a text file by using the Google Cloud Speech-to-Text API to try pattern matching on the text file, in order to determine the presence or absence of voice phishing. Pattern matching for text files can be performed more quickly and accurately using the Knuth–Morris–Pratt (KMP) algorithm, and the pattern detection time is negligibly short. The KMP algorithm is a string-matching algorithm that,unlike brute force searching, remembers failed search information and improves the speed of searching patterns used in the next search, as shown in Fig. 1 [26].

This paper looks at how to prevent voice phishing, a major social problem, in advance. By using the pattern matching KMP algorithm, the results of pattern detection can be known in real time. As such, this paper proposesa system that can recognize and block voice phishing in real time.

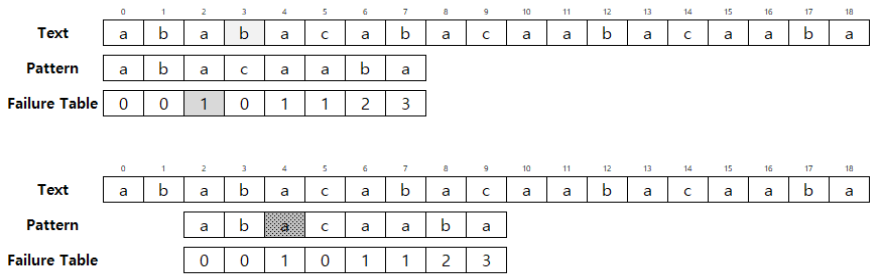

Fig. 1. Key word matching detection.

3. Proposed Voice Phishing Detection System

This study was conducted according to the process shown in Fig. 2, and the collected voice phishing sound source files were converted into text using a voice recognition API. Text classification was performed and accuracy was compared using the LSA method and the Doc2Vec method. Text data were preprocessed by extracting nouns, and each word was embedded using TF-IDF and Doc2Vec. The similarity between the documents was calculated through the evaluation classes of the LSA and Doc2Vec for the embedding results, and the accuracy of the voice phishing judgment was calculated based on the pre-classified voice phishing text and the plain currency text label [27, 28].

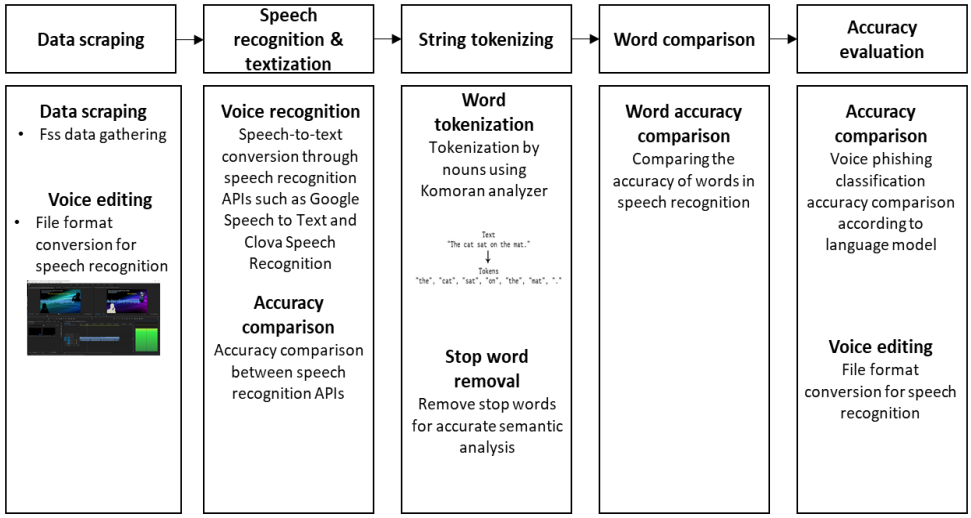

Fig. 2. Experimental process.

3.1 Voice Recognition

This study used the Google Cloud Speech API and Naver'sCLOVA Speech Recognition API for speech recognition, andcompared their performance. For this purpose, a transcription script was created for the voice phishing sound source file. The transcription script was written in such a way that the researcher could listento it directly, and if the voice was unclear or ambiguous during the writing process, the entire word was omitted. In the case of the Google Cloud Speech API, there are several restrictions for the speech recognition of voice files used as inputs, so the file format was converted to a voice file size of 1 MB or less, a sampling rate of 8,000 Hz, and a sound source length of less than 1 minute, in order to perform voice recognition. In the case of the CLOVA Speech Recognition API, there was no capacity limitation, but the same file format was used as input for comparison.

3.2 Text Tokenizing

Unlike English, which "tokenizes" words into units of spaces when performing text classification, Korean does not tokenize words in consideration of "words", which are units of spaces. It has the characteristics of an agglutinative language, so in Korean tokenization, the smallest unit of speech, the morpheme, must be considered [29]. One voice phishing datum generated through voice recognition was defined as one document, and the accuracy of the document classification model was compared by

looking atsimilar degrees of differencebetween the voice phishing document and the general call document.

4. Experiments

4.1 Voice Phishing Dataset

In this study, voice phishing data were obtained based on the voice phishing public sound source provided by the Korean Financial Supervisory Service. The sound source data are produced for educational and reporting purposes and provided in a video format, and the pseudonymization process for victims is also provided. In the case of voice phishing data, both institutional impersonation type and borrower type data were collected and analyzed.

Two hundred eighty-six voice phishing sound sources (.wav) were separated into 82 video files (.mpeg), and sound sources within 1 minute of the initial call were created to compare the proposed model's performance in detecting voice phishing in real time.

4.2 Textualization through Speech Recognition

For voice recognition, this study used the Google Cloud Speech API and Naver'sCLOVA SpeechRecognition API, and generated a script recorded manually from a voice phishing sound source file in order to compare their voice recognition performance.

The most basic indicator used when evaluating a classification model is accuracy. However, since accuracy does not consider the data imbalance of each category, a proper model performance evaluation may not be possible [30]. Therefore, in this study, accuracy, recall, precision, and the F1 score were used as evaluation indicators to consider category imbalance in addition to accuracy.

The average accuracy of speech recognition for voice phishing sound sources was found to be 56.5% for Google and 41.9% for Naver, and it was found that about 50% of the words that appeared were correctly recognized. When voice recognition was performed on the general call data, the average accuracy values were 42.0% for Google and 36.4% for Naver.

4.3 String Tokenizing

The voice phishing data used in this study are composed of sentences containing non-standard language, slang, etc.However, there is a problem in that the accuracy of tokenization is relatively low when using a general natural language processing library because the spacing is not well observed [31]. This is because an out-of-vocabulary problem that does not properly recognize words that did not appear in the learning data occurs when using the analyzer. Therefore, we used KoNLPy, a Python library for Korean natural language processing that is effectiveterms of data tokenization [32].

4.4 Sentence Embedding

To compare the accuracy of sentence-level embedding methods, embedding was performed in two ways: LSA and Doc2Vec.In the case of LSA, it created a TF-IDF matrix by treating the voice phishing sound source as a document and used the LSA model to implement document embedding. In this case, embedding was carried out in a 100-dimensional space using Doc2Vec in the Gensim module of Python, for which the top 10 document scores with high embedding scores are shown in Fig. 3.

4.5 Comprehensive Evaluation of Performance

The results of the comparison of the similarity of the documents based on cosine similarity, as shown in Figs. 4 and 5, clearly show that accuracy and precision were higher in the results embedded in the Doc2Vec method than those embedded in the LSA method.

```
[16]: from models.sent_eval import LSAEvaluator
      model = LSAEvaluator("./09.merge_data/d.clova_merge_one/clova_lsa.vecs")
      model.most_similar(doc_id=100)

[16]: ['titles100',
       [('titles93', 0.796999510401343),
        ('titles161', 0.7024041834509547),
        ('titles73', 0.5100190216908906),
        ('titles74', 0.3765001541428824),
        ('titles274', 0.3376625393969108),
        ('titles67', 0.29376153023089757),
        ('titles95', 0.28513796914082445),
        ('titles35', 0.2827110676757265),
        ('titles99', 0.27784859577815135),
        ('titles81', 0.27001507940584357)]]
```

Fig. 3. Document similarity.

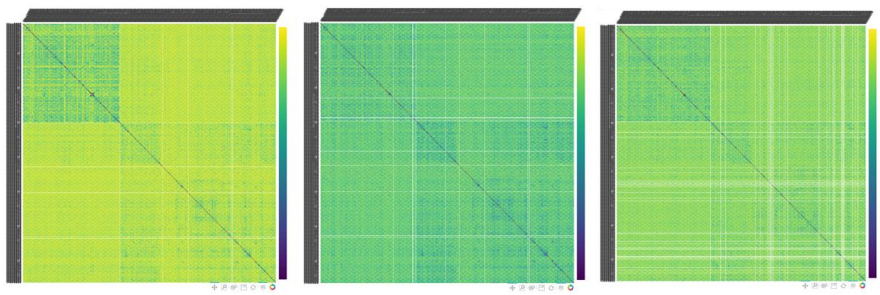

Fig. 4. LSA embedding result.

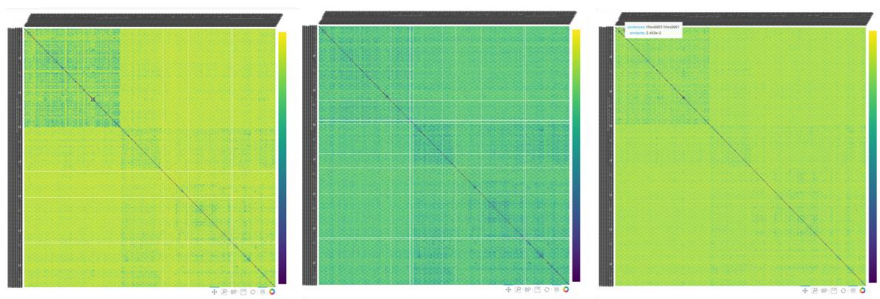

Fig. 5. Doc2Vec embedding result.

Table 3. Composite performance indicator

Method	Accuracy	Recall	Precision	F1
LSA, Handwritten script	0.57	0.70	0.58	0.63
LSA, CLOVA voice recognition	0.55	0.74	0.55	0.63
LSA, Google voice recognition	0.53	0.72	0.55	0.62
Doc2Vec, Handwritten script	0.60	0.52	0.66	0.58
Doc2Vec, CLOVA voice recognition	0.61	0.71	0.77	0.74

Doc2Vec, Google voice recognition	0.57	0.79	0.67	0.73

The specific evaluation scores are shown in Table 3, where a comprehensive performance indicator is derived by reflecting specific figures in the embedded results, asshown in Figs. 4 and 5. Each of the highest score of accuracy, precision, and F1 in six methods shown in Table 3 is 0.61, 0.77, and 0.74 in the method of “Doc2Vec, CLOVA voice recognition”. The highest score of recall in six methods shown in Table 3 is 0.74 in the method of “LSA, CLOVA voice recognition”.

4.6 Results

First, the sound source file obtained from the Financial Supervisory Service website was processed and converted to enable speech recognition, which was performed on the converted data to compare the accuracy of the handwritten script and the speech recognition results. Later, through the tokenization of words, the words that appearedfrequently in voice phishing documents were identified, and, in the process, the rate of recognition of a specific word was low, and constraintsdesigned to increase the level of accuracy were derived by excluding those wordswhen recognizing voice phishing. Thereafter, embedding based on frequency assumptions and distribution assumptions were generated using the LSA method and the Doc2Vec library in order to compare similarities. In order to evaluate the performance of the proposed methodology, the performance of each category was measured by comparing a voice phishing sound source and a general sound source. From the proposed methodology, it was confirmed that the Doc2Vec embedding method and the similarity determination method showed better performance. Through this, the proposed methodology confirmed that voice phishing can be judged by document data that are textualized by voice recognition for voice phishing sound sources.

5. Conclusion

This study proposes a methodology that compares the accuracy of classification of different methods of word embedding when new voice phishing data are generated by calculating the similarity of documents based on text as a result of speech recognition for voice phishing sound sources. The proposed methodology confirmed that voice phishing can be judged by document data that are textualized by voice recognition for voice phishing sound sources. This paper adopted a technical approach to the prevention of voice phishingand,unlike studies that approached the matter from a social engineering or technical perspective,derivedthe algorithms, speech recognition, accuracy analysis of natural language processing, constraints, etc. Its effectiveness was proved by conducting an analysis with phishing sound source data.

As regards the significance of this study, first, the model that classifies voice phishing by converting speech to text through self-language processing showed significant accuracy in voice phishing classification compared to the results of text matching or sound source analysis. Second, when textualizing Korean voice phishing voice data through voice recognition, the rate of recognition of a specific word deteriorates regardless of the type of voice recognition API, making it necessary to build a model that considers these words. In the voice classification of voice phishing through sentence embedding, the LSA method shows a significant difference from the model based on the handwritten listening script and the model based on the voice recognition text, making it necessary to prepare a classification model considering the difference.

As regards the limitations of the study, first, some 280 voice phishing source data were used to create model, so the size of the sample group is small, making it difficult to generalize about all voice phishing calls. Since the general call data collected for the purpose of comparison is personally collected data, the result obtained by comparing the data and voice phishing voice has a limitation in that it is difficult

to generalize it as actual voice phishing classification accuracy. Second, to protect personal information from voice phishing original voice data, it is difficult to understand the effect of voice, including the personal information of the victim, on the classification model analysis by using the data with some information deleted. Third, since the accuracy of each morpheme analysis library used for text tokenization was not compared, and a classification model was prepared by extracting only nouns from among morphemes, furtherresearch including a comparative analysis is required. As regards the direction of future work, in order to increase the accuracy when converting telephone speech to text, it will be necessary tocompare accuracy using various morpheme analysis libraries and to tokenize morphemes other than nouns in order to conduct additional comparative analysis studies.

Acknowledgements

Not applicable.

Author's Contributions

JWK proposed the main conception of the work and designed it; GWH analyzed and discussed the results; and HC supervised the entire work. All of the authors read and approved the final manuscript.

Funding

This paper was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008703).

Competing Interests

The individual contributions of the authors to the manuscript should be specified in this section.

References

[1] E. S. Jeong,"Study on intelligence model (AI) for detection of abnormal signs in electronic banking," M.S. thesis, Department of Information Security, Korea University, Seoul, Korea, 2018.

[2] G. C. Lee, *Korean Embedding.* Seoul, Korea: Acorn Publishing, 2019.

[3] C. H. Lim, "The need for active judicial relief against electronic financial fraud," *Law Journal*, vol. 65, pp. 257-282, 2019.

[4] S. J. Choi and J. B. Kim, "Comparison analysis of speech recognition open APIs' accuracy," *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, vol. 7, no. 8, pp. 411-418, 2017.

[5] J. C. W. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, and M. Aloqaily, "Privacy preserving multi-objective sanitization model in 6G IoT environments," *IEEE Internet of Things Journal*, 2020. https://doi.org/10.1109/JIOT.2020.3032896

[6] J. Salminen, M. Hopf, S. A. Chowdhury, S. G. Jung, H. Almerekhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Human-centric Computing and Information Sciences*, vol. 10, article no. 1, 2020. https://doi.org/10.1186/s13673-019-0205-6

[7] K. Choi, J. Lee, and Y. Chun, "Voice phishing fraud and its modus operandi," *Security Journal*, vol. 30, no. 2, pp. 454-466, 2017.

[8] J. Boehm, J. Kaplan, and N. Sportsman, "Cybersecurity's dual mission during the coronavirus crisis," 2020 [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Risk/Our%20Insights/Cybersecuritys%20dual%20mission%20during%20the%20coronavirus%20crisis/Cybersecuritys-dual-mission-during-the-coronavirus-crisis.pdf.

[9] J. H. Chang and K. H. Lee, "Voice phishing detection technique based on minimum classification error method incorporating codec parameters," *IET Signal Processing*, vol. 4, no. 5, pp. 502-509, 2010.
[10] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Transactions on Cybernetics*, vol. 45, no. 2, pp. 242-252, 2015.
[11] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proceedings of the 14th International Conference on Machine Learning*, San Francisco, CA, 1997,pp. 412-420.
[12] T. Joachims, "Text categorization with Support Vector Machines: learning with many relevant features," in *Machine Learning: ECML-98*. Berlin, Germany: Springer, 1998. pp. 137-142.
[13] A. Stolcke, "SRILM:an extensible language modeling toolkit," in *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP)*,Denver, CO, 2002, pp. 901-904.
[14] B. Li, Z. Zhao, T. Liu, P. Wang, and X. Du, "Weighted neural bag-of-n-grams model: new baselines for text classification," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*,Osaka, Japan, 2016, pp. 1591-1600.
[15] Y. Zhang and Z. Rao, "n-BiLSTM: BiLSTM with n-gram features for text classification,"in*Proceedings of 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*,Chongqing, China, 2020, pp. 1056-1059,
[16] B. Tang, S. Kay, and H. He, "Toward optimal feature selection in naive bayes for text categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2508-2521, 2016.
[17] F. Rousseau, E. Kiagias, and M. Vazirgiannis, "Text categorization as a graph classification problem," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China, 2015, pp. 1702-1712.
[18] S. T. Dumais,"Latent semantic analysis," *Annual Review of Information Science and Technology*, vol. 38, no. 1, pp. 188-230, 2004.
[19] N. Rizun, P. Kaplanski, and Y. Taranenko, "Development and research of the text messages semantic clustering methodology," in *Proceedings of 2016 Third European Network Intelligence Conference (ENIC)*, Wroclaw, Poland, 2016, pp. 180-187.
[20] S. Kim, H. Park, and J. Lee, "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis," *Expert Systems with Applications*, vol. 152, article no. 113401, 2020.
[21] Y. Goldberg and O. Levy, "word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method," 2014 [Online]. Available: https://arxiv.org/abs/1402.3722.
[22] A. Khatua, A. Khatua, and E. Cambria, "A tale of two epidemics: contextual Word2Vec for classifying twitter streams during outbreaks," *Information Processing & Management*, vol. 56, no. 1, pp. 247-257, 2019.
[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013 [Online]. Available: https://arxiv.org/abs/1301.3781.
[24] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," *Information Sciences*, vol. 477, pp. 15-29, 2019.
[25] M. C. Santillan and A. P. Azcarraga, "Poem generation using transformers and Doc2Vec embeddings," in *Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 2020, pp. 1-7.
[26] A. Rasool and N. Khare, "Parallelization of KMP string matching algorithm on different SIMD architectures: multi-core and GPGPU's," *International Journal of Computer Applications*, vol. 49, no. 11, pp. 26-28, 2012.
[27] S. A. Khanam, F. Liu, and Y. P. P. Chen, "Comprehensive structured knowledge base system construction with natural language presentation," *Human-centric Computing and Information Sciences*, vol. 9, article no. 23, 2019. https://doi.org/10.1186/s13673-019-0184-7
[28] J. C. W. Lin, Y. Shao, Y. Djenouri, and U. Yun, "ASRNN: A recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, vol. 212, article no. 106548, 2021.

[29] C. Kim, K. Kim, and S. Reddy Indurthi, "Small energy masking for improved neural network training for end-to-end speech recognition," in *Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 7684-7688.

[30] R. Benlamri and X. Zhang, "Context-aware recommender for mobile learners," *Human-centric Computing and Information Sciences*, vol. 4, article no. 12, 2014. https://doi.org/10.1186/s13673-014-0012-z

[31] X. Li, C. Yao, F. Fan, and X. Yu, "A text similarity measurement method based on singular value decomposition and semantic relevance," *Journal of Information Processing Systems*, vol. 13, no. 4, pp. 863-875, 2017.

[32] H. W. Seo, H. Kwon, M. A. Cheon, and J. H. Kim, "Bilingual multiword expression alignment by constituent-based similarity score," *Journal of Information Processing Systems*, vol. 12, no. 3, pp. 455-467, 2016.