Computer Science and Information Systems 21(2):663–683 https://doi.org/10.2298/CSIS230822012W

Multi-language IoT Information Security Standard Item Matching based on Deep Learning ⋆

Yu-Chi Wei[1], Yu-Chun Chang[2], and Wei-Chen Wu[3]

[1] National Taipei University of Technology
Taipei, Taiwan
vickrey@mail.ntut.edt.tw
[2] National Taipei University of Technology
Taipei, Taiwan
t109ab8013@ntut.org.tw
[3] Department of Finance, National Taipei University of Business
Taipei, Taiwan
weichen@ntub.edu.tw

Abstract. In the realm of IoT information security and other domains, various information security standards exist, such as the IEC 62443 series standards published by the International Electrotechnical Commission and ISO/IEC 27001 by the International Organization for Standardization. Business organizations are striving to improve and protect their operations through the implementation and study of these information security standards. However, comparing or pinpointing applicable control measures is becoming increasingly labor-intensive and prone to errors or deviations, especially given the plethora of information standards available. Identifying specific control measures scattered across different information security standards is gradually becoming an important issue. In this research, we utilise a range of domestic and international information security standards as the foundation, employing text mining and deep learning methods to map the similar parts of control measures between standards, thereby enhancing the efficiency of comparison tasks and allowing human resources to be allocated to more pertinent issues.

Keywords: Information Security, Information Security Standards, IoT Security, Text mining, Deep Learning.

1. Introduction

With the proliferation of Internet of Things (IoTs) technologies, everyday life has become increasingly digitized. IoT devices have a wide range of practical applications, whether in office environments, transportation, financial transactions, healthcare, or even in standard household smart appliances [22]. Broadly speaking, any device that can connect to the internet falls into this category, from those with basic network functionality to those combining various sensor devices, specialized software, or even capable of receiving and transmitting data from other complex IoT devices. The advent of IoT and the digital economy is a double-edged sword, on one hand making our lives more convenient to some extent, but on the other hand, escalating the information security threats associated with IoT devices and applications.

⋆ An extended version of The 12th Frontier Computing Conference/FC2022 paper

In the current era where the Internet of Things (IoT) is burgeoning and the interconnection of all things is becoming a trend, the potential risks behind its applications warrant our deep reflection and assessment. Revising new standards is a time-consuming and labour-intensive project, requiring information security professionals to reference, organise, and summarise the contents of various different standards. In this research, based on textual data exploration, existing international IoT standards are automatically preprocessed into numerous features, and then trained using deep learning models. This enables the automatic analysis of existing standards' information security requirements and their alignment with those of other international IoT information security standards. Finally, members of the standards drafting unit can directly refer to and assess whether the automatically generated corresponding results are suitable for use, thus saving a substantial amount of labour and time costs.

This research aims to utilise text mining to automatically translate and reference various existing international IoT standards. After textual preprocessing, these standards are trained using machine learning and deep learning models. The objective is to segment and automatically analyse the information security requirements of the existing standards, matching them with the requirements listed in other international IoT security standards. This not only assists the Mobile Application Security Alliance in continually updating IoT security verification standards, but also allows for the practical examination of whether domestic information security standard-setting processes comply with international IoT security standards. This study uses both domestic and international information security standard content as its dataset, with the capability to swiftly identify similar content. Furthermore, the content is not limited to being in the same language, and the overall output process can be finely tuned based on the input dataset to achieve the best matching results. This application is not limited to comparing and analysing the content of information security standards alone. It can also be based on other existing data and literature to explore and analyse their similarities, providing a reference for researchers looking to implement text processing, text analysis, machine learning, deep learning, and information security standards in their workflow.

2. Related Works

2.1. IEC 62443 Standards

IEC 62443 is a series of international standards for Industrial communication networks - IT security for networks and systems, which contains a series of technical procedures for the security of control systems, and the standard classified the user roles into operator, integrator and manufacturer, designs risks and potential problems for each role to help users of the standard to design and evaluate their own industrial automation systems and improve network security. The IEC 62443 series of standards is divided into four parts. The first part includes terminology and explanations of concepts related to automated industrial control systems, as well as examples of their use; the second part describes the security planning, operation, and management of the structure of industrial automation and control systems; the third part details technologies related to information security, information security risk assessments, and other definitions concerning information security; the fourth part focuses on the description of various security requirements, including

the product security development lifecycle, components, and technologies. In the process of developing the Mobile Application Security Alliance IoT security certification series, IEC 62443 Part 4-2 [8] (IEC 62443-4-2) is also one of the key reference standard and will be introduced separately in subsequent sections. The table below, Table 1, shows the structure and orientation of each content of the ”IEC 62443” series of standards.

Table 1. The list of IEC 62443 series of standards

Standard structure	Parts	Standard content
General: Defines the standard concepts, models, terminology interpretation and examples, etc.	1-1 TS	Concepts and models
	1-2 TR	Master glossary of terms and abbreviations
	1-3	System security compliance metrics
	1-4	IACS security life cycle and use-cases
Policies and Procedures: Provide defined system management requirements for IACS asset owners and services.	2-1 TS	Secure program requirements for IACS asset owners
	2-2	Security Protection Rating
	2-3 TR	Patch management in the IACS environment
	2-4 IS	Requirements for IACS service providers
	2-5 TR	Implementation guidance for IACS asset owners
System: Security risk assessment and security requirements defined for industrial control systems.	3-1	Security technologies for IACS
	3-2	Security risk assessment and system design
	3-3	System security requirements and security levels
Component: The safety product development process and component safety requirements as defined by the product supplier.	4-1 IS	Secure product development lifecycle requirements
	4-2	Technical security requirements for IACS components

2.2. OWASP Top 10

OWASP, known as the Open Web Application Security Project, is an open, non-profit organization dedicated to helping governments and businesses improve web software security, tools, and technical documentation, as well as gain practical insight into the vulnerabilities and security of the information assets they use. Every few years, OWASP produces a list of the top 10 web application security vulnerabilities and provides some easy ways and directions to educate users on how to avoid these vulnerabilities. Table 2. below shows the ten web application security vulnerabilities pro-posed in ”OWASP Top 10:2021 [18]”.

Despite all the vulnerabilities presented in the OWASP Top 10 are carefully organized and filtered to the top ten most common web application security vulnerabilities of our time, there is still a ranking hierarchy among the vulnerabilities, and the higher the ranking, the more important the web application security vulnerability is in the current information environment.

Among the existing web application security vulnerabilities, there are several items that have appeared in the previous version of the ”OWASP Top 10”, but their ranking has been changed in response to the changing times and environment. For example, A01:

Table 2. The list of OWASP Top 10:2021

Vulnerabilities ID	Vulnerabilities Top 10 of application security ver.2021
A01:2021	Broken Access Control
A02:2021	Cryptographic Failures
A03:2021	Injection
A04:2021	Insecure Design
A05:2021	Security Misconfiguration
A06:2021	Vulnerable and Outdated Components
A07:2021	Identification and Authentication Failures
A08:2021	Software and Data Integrity Failures
A09:2021	Security Logging and Monitoring Failures
A10:2021	Server-Side Request Forgery

Access Control Failure in "OWASP Top 10:2021", which was ranked fifth in the previous version of OWASP Top 10:2017, was moved from fifth to first in the latest version. According to the officials, more than 90% of the applications they tested had a category access failure problem, and the number of occurrences was much higher than other vulnerability categories.

In addition to the ten most common security weaknesses of web applications, OWASP also has responded to the increasing use of APIs and Internet of Things devices in the industry, they presented the "OWASP API Security Top 10" and "OWASP IoT Top 10", which includes ten most common security vulnerabilities of network applications. Despite there is a newer version of "OWASP IoT Top 10", which is the version 2018, but overall and detailed information of "OWASP IoT Top 10:2014 [16]" is relatively more abundant than the 2018 version on the official OWASP website, more information is definitely more helpful for deep learning model to classify information security controls into similar categories, that was the main reason we chose to use "OWASP IoT Top 10:2014 [16]" instead of "OWASP IoT Top 10:2018 [17]". Table 3 below shows the list of top 10 security vulnerabilities of "OWASP IoT Top 10:2014".

Table 3. Introduction to the OWASP IoT Top 10:2014

Vulnerabilities ID	Vulnerabilities Top 10 of Internet of Things ver.2014
I01:2014	Insecure Web Interface
I02:2014	Insufficient Authentication/Authorization
I03:2014	Insecure Network Services
I04:2014	Lack of Transport Encryption
I05:2014	Privacy Concerns
I06:2014	Insecure Cloud Interface
I07:2014	Insecure Mobile Interface
I08:2014	Insufficient Security Configurability
I09:2014	Insecure Software/Firmware
I10:2014	Poor Physical Security

For the showcase of this study, we attempted to make IEC 62443-4-2 controls automatically classified into the closest of the ten specified "OWASP IoT Top 10" categories through text mining and deep learning methods, thus saving the time and cost required for manual comparison of information security standards.

2.3. Text Similarity Matching

Text similarity matching methods are becoming increasingly important in many applications. Existing methods often compute similarity based on shallow syntax or POS tagging or by comparing basic syntax similarity, generating vectors, and then inferring similarity from this set of vectors. However, due to the variability in natural language expression, these methods often struggle to predict actual semantic content and implications. To address these issues, researchers have attempted various approaches. At the beginning, using a lexicon to note the positions of words within sentences, forming a one-hot encoding vector representation. However, this method couldn't link related words. Mikolov et al. citeMikolov2013 tried combining neural networks in their research. Turian et al. [23] tried using pre-trained word representations in conjunction with supervised learning methods as extra features, which showed significant improvements over traditional word embedding methods. These methods evolved into larger frameworks, like sentence embedding [10] or paragraph embedding [11]. Matthew et al. [19] extracted context-sensitive features from language models, integrated these features into training for specific tasks, and gradually began to understand the variability and actual semantic content in language expressions.

Researchers also considered the context and situation of sentences. The Skip-gram model [5] is a renowned method that trains and identifies using the context of target words. WordNet [4] is a network primarily focused on the "word-semantics" in English, storing the structures and potential relationships between words, quantifying the semantic relationship between two different words. ConceptNet [12] uses a dictionary-based embedding model, aligning with the hierarchical structure of predefined words in WordNet, defining various relationships between words. Emrah [7] proposed a method focused on calculating sentence similarity without using machine learning, relying on dependency parsers and lexical embedding models, achieving results better than most traditional methods.

In research on machine learning for text similarity analysis and comparison, Ji and Eisenstein [9] introduced a supervised machine learning method that measures semantic similarity between sentences using a discriminative term or proper noun, in conjunction with a set weighting index, giving higher importance to certain features, then computing sentence similarity. The authors claimed their new method outperforms the widely-used TF-IDF weighting method. Mohamed and Oussalah [15] proposed a similarity calculation method that uses WordNet to obtain dependency relationships for words which based on instances extracted from Wikipedia and normalized Google distance. The normalized Google distance calculates the hit count returned for a set of keywords using the Google search engine. Hassan [6] proposed a method based on Wikipedia's content for context determination, called Salient Semantic Analysis (SSA). Mihalcea et al. [13] combined corpus-based semantic similarity with knowledge-based semantic similarity, using data from WordNet and the British National Corpus, which reduced the error rate compared to traditional methods. Wang et al. [24] focused on the similar and dissimilar parts of

sentences. They constructed a similarity matrix and corresponding vectors for each word meaning, decomposing the resulting matching vectors to identify similar and dissimilar parts, eventually using matrix decomposition to extract sentence vectors to compute sentence similarity.

BERT [2], introduced by Google's AI team in 2018, used BooksCorpus and over 800 million entries and data from Wikipedia for pre-training. The operation is divided into two stages: pre-training and fine-tuning. In the pre-training phase, there are two training methods: Masked LM and Next Sentence Prediction. Then in the fine-tuning phase, the model is adjusted based on specific tasks. BERT performs well in sentence classification, tagging, and text classification. However, Reimers and Gurevych [20] found that while BERT and RoBERTa achieve effects in many sentence regression tasks, such as text semantic similarity, they need to input two sentences to be compared into the model repeatedly until the closest two sentences are found. The excessive computational cost makes BERT unsuitable for semantic similarity searches. Hence, they introduced SBERT (Sentence-BERT). SBERT, unlike BERT, which repeatedly attempts to combine two sentences, calculates the similarity distance between two sentences directly by matching their word embedding representations, significantly reducing computation. This model also achieves good results in some STS and transfer learning tasks.

3. Research Methodology

3.1. Data Pre-processing

In this study, we use python and jupyter notebook as the test environment. In the data pre-processing progress, we first need to retrieve the contents of the information security standard, and split the contents into each column, including its control number, control name and control description as a spreadsheet. After this, the contents of the information security standard form are stored in memory and ready to go.

These manually retrieved control contents in the spreadsheet does not require data pre-processing, they can simply import into the deep learning model in their original format for training. The pre-trained models provided by SBERT [21] are already trained from various types of datasets, familiar with the original word patterns, so there is no need to perform steps such as words and sentences segmentation, word lemmatization, stemming or other data pre-process methods you can find in other NLTK tasks to filter the features.

As shown in the figure above, the following figure is a screenshot of the jupyter notebook after importing and reading the information security standard content into the spreadsheet. This experiment uses IEC 62443-4-2 [8] content as the training set, and tries to classify the content of the controls in each of the ten categories of "OWASP IoT Top 10:2014 [16]" as the test set.

It is also possible to match similar contents between different language information security standards. In the data pre-processing state, while keeping the unique control id number field legible, translation modules can be used to translate control descriptions into the specified language and then perform a similarity comparison exercise with other information security standards. Usually, it is better to translate other languages into English and perform similarity matching between the two standards using English as the common language, because most of the pre-training data for deep learning models are trained from English data as shown in the Fig. 2 below.

```
In [17]: train_blist[0]

Out[17]: 'Components shall provide the capability to identify and authenticate all human
         users according to IEC 62443-3-3 SR 1.1 on all interfaces capable of human user
         access. This capability shall enforce such identification and authentication on
         all interfaces that provide human user access to the component to support segre
         gation of duties and least privilege in accordance with applicable security pol
         icies and procedures. This capability may be provided locally by the component
         or by integration into a system level identification and authentication syste
         m.'

In [18]: test_blist[0]

Out[18]: 'Default passwords and ideally default usernames to be changed during initial s
         etup'
```

Fig. 1. A schematic diagram of train/test data content

	影像監控系統標準_zh	影像監控系統標準_translated
0	產品預設不應透過實體介面存取產品作業系統之除錯模式。 若需經實體介面存取，則應通過身分鑑別作...	Product presets should not be removed by the p...
1	產品應具有實體埠插拔操作記錄功能。	The product should have the function of the ph...
2	產品應具備相關警示功能於實體操作發生斷訊時。	The product should have the relevant warning f...
3	產品外部不應有徒手即可還原預設通行碼的功能。	The outside of the product should not have the...
4	產品應支援安全啟動(Secure Boot)功能，不應以未經授權的韌體、驅動程式及作業系統執...	The product should support the Security Boot f...
5	產品之作業系統與網路服務，不應存在美國國家弱點資料庫所公開的常見弱點與漏洞資料，且漏洞評鑑系...	The operating system and network services of t...
6	產品之作業系統與網路服務，不應存在美國國家弱點資料庫所公開的常見弱點與漏洞資料，且漏洞評鑑系...	The operating system and network services of t...
7	產品開啟之網路服務應為廠商提供必要服務之所需，防止產品因啟用網路介面而被侵入的可能性，且廠商...	The online service of the product opening shou...
8	產品所收集之遙測資料應告知使用者，且未告知之遙測資料不應被收集。	The remote measurement data collected by the p...
9	韌體應具備更新機制。	The ligament should have a update mechanism.
10	產品若支援離線手動更新，則更新檔案應加密保護以確保機密性，且應採用NIST SP 800-1...	If the product supports offline manual update,...
11	產品若支援線上更新，其更新路徑應通過安全通道，且安全通道版本應符合「附錄 A」的要求，同時金...	If the product supports online update, its upd...
12	產品應具備驗證韌體之完整性及真確性的功能。	The product should have the function of verify...
13	產品應具備備援更新功能，即發生更新失敗時，系統能回復至更新前之狀態。	The product should have the renewal function, ...
14	產品所儲存的敏感性資料，應僅由獲授權個體存取。	The sensitive data stored in the product shoul...
15	產品所儲存之身分鑑別因子、加解密用之金鑰(不含非對稱加密用之公鑰)不應明文儲存，而保護資料的...	The identity of the product's identity identit...
16	產品應提出金鑰管理程序，以確保金鑰管理的品質。	The product should propose the golden key mana...

Fig. 2. A schematic diagram illustrating the successful prediction of control items between two different standards

In the Fig. 2 above, we used a local IoT security standard for this showcase, which is "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]" from Mobile Application Security Alliance, which is an IoT product certification alliance dedicated to the promotion of domestic IoT information security in Taiwan. According to the figure, any standards in different languages can be translated into English by the translation module and then start the comparison process of information security standards directly, this allows the process to be able to compile information security standards in different languages without any limitation due to language.

3.2. Model Training

BERT [1], the abbreviation of Pre-training of Deep Bidirectional Transformers for Language Understanding, which is already highly characterized by the endless training data based on the Google search engine, so that BERT only needs to specify the form of its output data, and then fine-tune it according to the task, finally, it can be used for various common natural language processing tasks.

But Reimers and Gurevych found that although both BERT and RoBERTa achieve some good results in many sentences regression tasks, such as textual semantic similarity, they both need to pass both sentences to be compared into the model and repeat this process until the two most similar sentences are found. This is a very costly process, especially when the data is large. According to this, they considered BERT is not suitable for the task of semantic similarity search because of the limitation of the algorithm, so they proposed SBERT [21] (Sentence-BERT), which does not need to try to combine two sentences repeatedly like BERT, but by directly matching and calculating the words similarity distance of two sentences using word's embed-ding representations, which greatly reduces the computational effort and achieves very good results in some STS and migration learning tasks.

The deep learning method SBERT provides a number of pre-training models, which allow users to train their own research data directly to make further predictions. In the official guidance document of SBERT, 13 pre-training models are provided. The 13 pre-training models are listed with their performance of sentence embeddings, performance of semantic search, average overall performance, running speed and model size, so that users can select them according to their task requirements. The five models with the best performance based on the above five indicators were shown as Table 4.

In this study, the best average overall performance one: *all-mpnet-base-v2*, were selected, which was an all-round model tuned for many use-cases, trained on a large and diverse dataset of over 1 billion training pairs.

As shown in the figure above, IEC 62443-4-2 controls were successfully classified by deep learning models into the ten categories of OWASP IoT Top 10:2014. Matching similar contents including controls or descriptions between several information security standards, which often requires a lot of labor and time, but this study showed that it is totally possible to quickly generate similarity comparison results between certain information security standards by using text mining and deep learning methods. It can also be said that this study, corresponding contents between information security standards and standards is also one of the typical NLP tasks, i.e., the application of semantic textual similarity tasks.

Table 4. Comparison of SBERT best performance pre-trained models

Model name	Performance of sentence embeddings	Performance of semantic search	Average overall performance	Encoding speed	Model size
all-mpnet-base-v2	69.57	57.02	63.30	2800	420 MB
multi-qa-mpnet-base-dot-v1	66.76	57.60	62.18	2800	420 MB
distiluse-base-multilingual -cased-v2	60.18	27.35	43.77	4000	480 MB
paraphrase-MiniLM-L3-v2	62.29	39.19	50.74	19000	61 MB
paraphrase-multilingual-mpnet-base-v2	65.83	41.68	53.75	2500	970 MB

	IEC 62443 Controls	OWASP Top 10:2014 Category	Distance
0	5.3 CR 1.1	I5	0.515
1	5.4 CR 1.2	I5	0.519
2	5.5 CR 1.3	I5	0.385
3	5.6 CR 1.4	I5	0.416
4	5.7 CR 1.5	I2	0.469
...	...	...	...
83	15.9 NDR 3.12	I5	0.575
84	15.10 NDR 3.13	I5	0.580
85	15.11 NDR 3.14	I10	0.528
86	15.12 NDR 5.2	I3	0.455
87	15.13 NDR 5.3	I3	0.530

Fig. 3. A comparative schematic illustrating the distance between control measures across different standards

3.3. Evaluation Methodology

In section 3.2, we have demonstrated that it is possible to perform similarity comparisons between information security standards using deep learning methods. But, how was the predictive accuracy? To find out the predictive accuracy of the model, first, a reference answer that cross-validates the model prediction results is necessary. For example, a table which providing an official mapping of the controls of a standard itself to the controls of another standard, such as a table which maps IEC 62443-4-2 [8] controls to EN 303-645 [3] controls. But unfortunately, no such mapping table is provided in the official documents of these two parties.

For this reason, we use the official mapping of Appendix D of IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14], which is a Taiwanese IoT information security standard focused on image monitoring systems, includes a mapping table to the OWASP IoT Top 10:2014 [16], these two information security standard have built a explicit relations between their controls, which allows this study to use the information in this table as a reference for the accuracy of automated comparisons with deep learning models. Fig. 4 below shows a screenshot of the controls in Appendix D of the standard "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements" against each standard specification.

附錄 D

(參考)

技術要求事項與各標準規範對照表

Control's serial number of the standard

表 D.1 技術要求事項與各標準規範對照表

技術要求	OWASP 對應項目(4)	ANSI/UL 2900-1 對應項目	ONVIF 對應項目(19-20)
5.1.1.1	I10： Poor Physical Security Ensuring only required external ports such as USB are required for the product to function. Ensuring the product has the ability to limit administrative capabilities.		-

Fig. 4. A screenshot of the standard " IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements, Appendix D " against OWASP IoT Top 10:2014

The reference answer mapping of the standard comparison is based on the ”IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements” standard, and the official mapping table of the standard to OWASP IoT Top 10:2014 in Appendix D of the standard as the reference answer. In other words, a total of 38 screened security items in the Image Monitor System Information Security Standard will actually be classified into the ten corresponding categories of ”OWASP IoT Top 10:2014”. Although each category of ”OWASP IoT Top 10:2014” has from 4 to 14 information security controls, it was found that it is difficult to match the information security control specified in the reference answer for information security standards from different sources. In addition to the difference in terminology between different standards, it is assumed that the accuracy of the wording of the original Chinese standard will be affected after translation. Therefore, in this section, we choose to convert the accuracy of the base standard information security control into reference values by whether they are correctly classified or not. Figure 5 below shows the schematic diagram of the two experimental approaches.

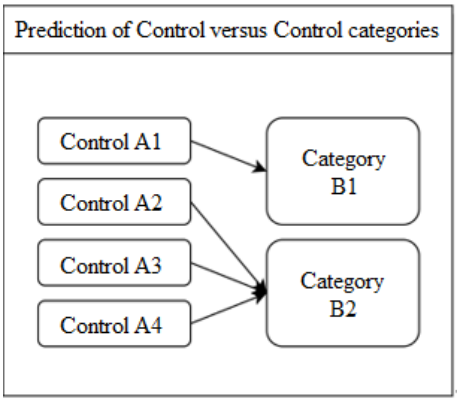

Fig. 5. A schematic diagram of controls versus control categories comparison

As shown in the Fig. 4, which shows that the control numbered 5.1.1.1 of ”IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]” can actually corresponded to category I10 of ”OWASP IoT top 10:2014 [16]”.

However, since the standard itself is written in Chinese, it needs to be translated into English and then fed into a deep learning model for comparison, so we have used the translation module mentioned in section 3.1 to automatically complete this task for us. After checking the table, a total of 38 filtered security controls in the ”IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements” will actually be classified into the ten corresponding categories of ”OWASP IoT Top 10:2014”.

4. Evaluation Results

4.1. Initial Evaluation Results

We used the five models with the best performance in Table 4. and *distiluse-base-multilingual-cased-v2*, which is a multilingual model that supports more than 50 different languages, and more balanced in the scores of the indicators, were selected and compared with the "OWASP IoT Top 10: 2014", and the following Table 5 shows the experimental results.

Table 5. Comparison of experimental results with different SBERT models

Exp. No.	Model	k=1	k=2
S1	all-mpnet-base-v2	61 %	68 %
S2	multi-qa-mpnet-base-dot-v1	50 %	66 %
S3	paraphrase-MiniLM-L3-v2	39 %	50 %
S4	distiluse-base-multilingual-cased-v2	68 %	68 %
S5	paraphrase-multilingual-mpnet-base-v2	50 %	74 %

In the above table, the k represents the prediction of the k most similar outcomes at the end of each prediction. In other words, when the model can output the least number of predictions, the more accurate it can hit the same category of predictions, which means that the model has a better performance on the task of matching information security standards. The number of successful hits is one of the important indicators of the effectiveness of the reference model for this task.

Under this condition, experiment number S1 and S4 have the best performance, which are *all-mpnet-base-v2* and multilingual model *distiluse-base-multilingual-cased-v2*, achieving 61% and 68% hit rate respectively under the restriction of k=1, and 68%, 74% hit rate respectively under the restriction of k=2, which means at least three quarters of controls in the standard were successfully predicted to the correct categories by the deep learning models.

In addition to the difference in terminology between different standards, the accuracy of the wording of the original standard will also be affected if it is translated, not to mention the fact that there are also controls or requirements that meet several OWASP IoT Top 10 categories after review and analysis, but the reference answer only has a given category and thus cannot be included. However, when it comes to the actual use for the information standards, even though they are for the same domain-oriented information security standards, there are some parts that are not similar. In practice, when an information security consultant is looking for controls or requirements that are suitable for a particular case, the items that are suitable for the case may be scattered in different information security standards, or different categories inside the same standard. Among those that are not successful, there must be some items that are not in the same category but have similar practical applications and application methods.

4.2. Discussion of Evaluation Results

In Section 3.2 of this paper, the five SBERT models that performed better on average were compared with the results of the comparison experiments between their scores provided in

the official guidance documents and the English standards. This means that nearly a quarter of the information security items are difficult to classify correctly by the model. The actual list of information security item numbers that were not predicted by each model shows that these unpredictable information security item numbers are specific numbers, as shown in Fig. 6 below shows the prediction status of each model for the specified corresponding security item at k=3. The dark squares indicate that the number was successfully predicted by the specified model, while the light squares indicate that the number was not successfully predicted by the specified model.

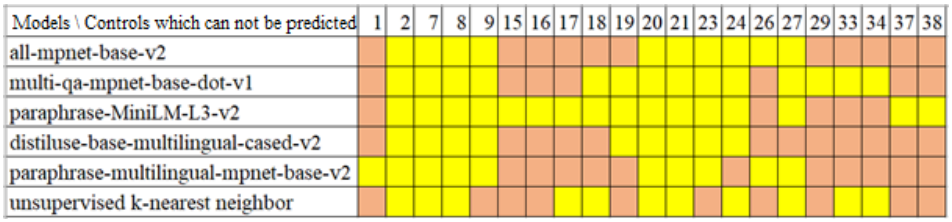

Fig. 6. Standard controls for which none of the plural models can be predicted

The distribution of the light-colored squares in the above figure shows that the information security items that cannot be successfully predicted by the specified models are very similar for the above five deep learning models, especially for Experiment Numbers. 2, 7, 8, 9, 20, 21, and 23. Experiment number 2, 7, 8, 9, 20, 21, 23, these experiment numbers correspond to the following items in the original standard: "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]".

Based on the above table, it can be inferred that it is more difficult for the deep learning models to classify the contents of information security controls into two categories, category 2 and category 8. When encountering the above information security controls in practice, both category 2 and category 8 will not be the first choice of the deep learning models, but other categories. In" OWASP IoT Top 10:2014" [16], category 2 is Insufficient Authentication/Authorization, which translates to unreliable authentication mechanism, and category 8 is Insufficient Security Configurability, which translates to unreliable security configuration.

From the evaluation results, the two deep learning models with the best prediction results, *all-mpnet-base-v2* and *distiluse-base-multilingual-cased-v2*, are the best predicted models for the translated "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]", which corresponds to the prediction of "OWASP IoT Top 10:2014", achieved 61% and 68% hit rate at k=1 respectively. The prediction results are shown in Table 7 below.

According to the above table, the prediction results can be easily classified into two categories, one is the category where Model 1, denoted as $M1$, and Model 2, denoted as $M2$, have the same prediction results for the information security sub-category, but both predict failure. Table 8 explores the potential causes of prediction errors in the classification results. In Experiment Number 2, the correct category should in Category 2, the

Table 6. List of standard controls that cannot be predicted by the majority of models.

Exp. No.	Control Number	Correct Category	Control Description
2	5.1.3.1	2	The product should not have the ability to restore the default pass code with your bare hands.
7	5.2.4.1	8	Sensitive data stored in the product shall be accessible only by authorized individuals.
8	5.2.4.2	8	The identity authentication factor and key for encryption and decryption (excluding the public key for asymmetric encryption) stored in the product should not be stored in clear text, and the data should be protected by the security functions approved by NIST SP 800-140C, CMVP Approved Security Functions.
9	5.2.4.4	8	Sensitive data should be stored in the security domain of the product, isolated from the normal operating environment.
20	5.3.3.1	8	The product should provide the user to turn on/off the WPS PIN function of ”Wi-Fi Protected Setup (WPS)” and its default value should be off.
21	5.3.3.2	8	By default, the Wi-Fi security mechanism should be ”Wi-Fi Protected Access (WPA)” and the version of Wi-Fi Protected Access should meet the requirements of Appendix C.
23	5.4.1.1	2	Before accessing the product resources, the identity identification mechanism with protection against retransmission attacks should be adopted.

Table 7. Prediction results of the two models for the specified standard controls

Exp.No.	Control	$M1$ prediction result	$M2$ prediction result
2	5.1.3.1	10	10
7	5.2.4.1	5	5
8	5.2.4.2	4	4
9	5.2.4.4	5	10
20	5.3.3.1	7	5
21	5.3.3.2	7	7
23	5.4.1.1	10	7

corresponding information security subdivision is that the product should not have the ability to restore the default passcode externally with bare hands. It should be the word ”external” that causes the deep learning model to predict this information security item as Category 10: Poor physical security. In Experiment Number 7, the corresponding information security breakdown is that sensitive information stored in the product should only be accessed by authorized individuals. In terms of this information security control, it is reasonable to predict to Category 5: privacy concerns because it also describes user privacy. In Experiment Number 8, The corresponding information security itemized content is: the identity authentication factor and key for encryption and decryption (excluding the public key for asymmetric encryption) stored in the product should not be stored in clear text, and the data protection method should be used with the security functions approved by NIST SP 800-140C, CMVP Approved Security Functions. In terms of this information security category, the prediction to Classification 4: Lack of Transport Encryption is reasonable because it contains key words in the field of encryption such as encryption and decryption, key and plaintext. In Experiment Number 21, for the information security control, the default security mechanism for Wi-Fi is ”Wi-Fi Protected Access (WPA)” and the version of Wi-Fi Protected Access should meet the requirements of Appendix C. In terms of this information security control, the predicted classification is Category 7: Insecure Mobile Interface, which is not accurate. It is guessed that in the pre-training data of the two models, Wi-Fi usually appears together with key words such as cell phone and mobile, so the models classified it as Category 7.

Table 8. The two models jointly classify the error causes of the false security item

No.	Control	Result ($M1\&M2$)	Correct category
2	5.1.3.1	10	2
7	5.2.4.1	5	8
8	5.2.4.2	4	8
21	5.3.3.2	7	8

The reason behind the inaccurate predicts are that different deep learning models have different prediction judgments for the same information security item, but the classification is basically similar to the former one: it is influenced by specific wording, or the information security item may apply to both plural ”OWASP IoT Top 10:2014” classification, resulting in its misclassification. The actual results of the respective predictions are listed for analysis, and the reasons for the wrong classification results are speculated in Table 9. In Experiment Number 9, the corresponding information security sub-section is: sensitive data should be stored in the security domain of the product, isolated from the normal operating environment. Model 1 predicts that Category 5: Privacy Concerns are reasonable, and sensitive data are indeed related to user privacy; Model 2 predicts that Category 10: Poor Physical Security is not reasonable, and the model presumes that the information security item is not related to physical security because of the terms ”operating environment”, ”isolation”, and ”security domain”. The model predicts that the information security item is related to the description of physical security because of the terms ”operating environment,” ”isolation,” and ”secure area. In Experiment Number 20,

the information security control is: the product should provide users to turn on/off the WPS PIN function of ”Wi-Fi Protected Setup (WPS)”, and the default value should be off. Model 1 predicts Category 7: Insecure Mobile Interface, which is a relatively inaccurate classification. It is guessed that in the pre-training data of both models, Wi-Fi usually appears together with key words such as cell phone and mobile, so the model classifies it as Category 7. After all, if Wi-Fi is automatically connected to public networks, it may cause user privacy leakage, which is a user privacy concern.

In Experiment Number 23, the corresponding information security control is: Before accessing product resources, identity authentication mechanism with protection against retransmission attacks should be used. Model 1 predicts a classification of 10: Poor Physical Security, which is inaccurate. Model 2 predicts a classification of 7: Insecure mobile interface, which is more reasonable than the prediction of Model 1, but not correct. In Experiment Number 21, for the information security control, the default security mechanism for Wi-Fi is ”Wi-Fi Protected Access (WPA)” and the version of Wi-Fi Protected Access should meet the requirements of Appendix C. In terms of this information security control, the predicted classification is Category 7: Insecure Mobile Interface, which is not accurate. It is guessed that in the pre-training data of the two models, Wi-Fi usually appears together with key words such as cell phone and mobile, so the models classified it as Category 7.

Table 9. The two models each classify the wrong security category of the error cause speculation

No	Control	Result($M1$)	Result($M2$)	Correct category
9	5.2.4.4	5	10	8
20	5.3.3.1	7	5	8
23	5.4.1.1	10	7	2
21	5.3.3.2	7	7	8

From Table 9 and the speculation on the failure of the prediction of the security category for which none of the plural models could be predicted, it is clear that at least half of the security categories that failed to be predicted may also apply to the plural "OWASP IoT Top 10:2014[16]" classification, plus the fact that in the standard "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]", the corresponding OWASP In Appendix D of the original Top 10:2014 mapping table, the security subcategory does not specify a mapping to another subcategory, even though the subcategory is similar for that security subcategory, resulting in model prediction failure. By actually viewing the table and the information security controls that failed for the seven security controls that could not be predicted by the plural model, if the predictions that were judged to be reasonable were categorized as correct predictions, with model 1 representing *all-mpnet-base-v2* and model 2 representing *distiluse-base-multilingual-cased-v2*, the two models The final revised prediction results for these seven information security controls are shown in Table 10 below.

Table 10. Results of the error analysis of the two models for the unpredictable security controls breakdown

Exp. No. / Ctrl. No.	2 / 5.1.3.1	7 / 5.2.4.1	8 / 5.2.4.2	9 / 5.2.4.4	20 / 5.3.3.1	21 / 5.3.3.2	23 / 5.4.1.1
Model 1	X	V	V	V	X	X	X
Model 2	X	V	V	X	V	X	X

According to the above table, model 1: *all-mpnet-base-v2* and model 2: *distiluse-base-multilingual-cased-v2* achieve 61% and 68% hit rate respectively for k=1. If the predictions with reasonable classification are classified as correct and recalculated, the hit rate will increase to 69% and 76%.

Finally, the experimental results proved that the use of deep learning models for fast and automated comparison of information security standard content has good accuracy and retains considerable room for improvement.

4.3. Final Evaluation Results

SBERT [21], as an enhanced version of BERT [1] for text similarity search task, provides a pre-training model with higher accuracy than the native pre-training model provided by BERT. Table below shows the best two models in SBERT, *all-mpnet-base-v2*

and *distiluse-base-multilingual-cased-v2*, with the same k=1, i.e., each information security sub-prediction only outputs one closest information security sub-prediction, and this output value is the only consideration for accuracy. Under the condition that the translated "IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements [14]" corresponds to the prediction of "OWASP IoT Top 10:2014"[16].

Table 11. Deep Learning Approach to Information Security Standard Prediction Implementation Results

Exp. No.	Model name	Predict accuracy	Predict / All
SS1	all-mpnet-base-v2	69%	26 / 38
SS2	distiluse-base-multilingual-cased-v2	76%	29 / 38

According to the above table, the better model, *distiluse-base-multilingual-cased-v2*, successfully predicted 26 of the 38 information security items with k=1, while the remaining items failed to be predicted by the plural model in sub section 4.2 of this study. In this study, we examined the seven information security items for which both models failed to predict, and confirmed that three of the items were also applicable to the plural "OWASP IoT Top 10:2014" classification, although they did not match the answers.

In spite of the information security standards targeting the same aspect, there will still be parts where they differ significantly from each other. In practice, when information security consultants are looking for suitable sub-items or control measures for a specific case, the relevant items may be spread across different categories. Among the items that don't align perfectly, there are bound to be some that, while not in the same category, are very similar in practical application and usage. This suggests that, in practical terms, using deep learning models for comparing information security standards has shown, from experimental results, to be not only faster but also fairly accurate. Its performance surpasses the implementation using traditional machine learning. In the future, this research will experiment with generative AI, attempting to produce more general terms related to the control items of different standard, and then apply the SBERT method for further experimentation to enhance the readiness of successful classification.

5. Conclusion

This study utilises the contents of multiple international information security standards and translated domestic standards as its dataset, possessing the ability to rapidly identify similar control items. The content is not restricted to a single language and demonstrates good predictive accuracy. The study also proposes an automated process, streamlining a workflow that would otherwise require significant labour to review and compare. Ultimately, this can serve as a reference for scholars wishing to conduct future research in text processing, text mining, deep learning, and information security standards.

Although this research has achieved commendable results in comparing similarities among different information security standards, there are still many areas that warrant

further exploration in the future. For instance, automated data processing procedures or the application of machine learning methods such as Few-Shot Learning for data with lower volume, greater diversity, and insufficient annotations. Additionally, the use of generative AI represents another avenue to explore. Some standards may feature different customary terminologies across various standards organisations or publishers. Generating more general terms related to control and then utilising the SBERT method for further experiments might enhance the accuracy of successful classifications.

Acknowledgments. This research was partially funded by National Science and Technology Council (NSTC 112-2221-E-027-067-).

References

1. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.N.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018), https://arxiv.org/abs/1810.04805
3. European Telecommunications Standards Institute: EN 303 645:CYBER; Cyber Security for Consumer Internet of Things: Baseline Requirements, v2.1.1 edn. (2020)
4. Fellbaum, C.: WordNet, pp. 231–243. Springer Netherlands, Dordrecht (2010), https://doi.org/10.1007/978-90-481-8847-5_10
5. Guthrie, D., Allison, B., Liu, W., Guthrie, L., Wilks, Y.: A closer look at skip-gram modelling. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06). European Language Resources Association (ELRA), Genoa, Italy (May 2006), http://www.lrec-conf.org/proceedings/lrec2006/pdf/357_pdf.pdf
6. Hassan, S.: Measuring semantic relatedness using salient encyclopedic concepts. Ph.D. thesis (2011), https://www.proquest.com/dissertations-theses/measuring-semantic-relatedness-using-salient/docview/1011651248/se-2
7. Inan, E.: Simit: a text similarity method using lexicon and dependency representations. New Generation Computing 38(3), 509–530 (2020)
8. International Electrotechnical Commission: Security for industrial automation and control systems - Part 4-2: Technical security requirements for IACS components. (2019)
9. Ji, Y., Eisenstein, J.: Discriminative improvements to distributional sentence similarity. In: Proceedings of the 2013 conference on empirical methods in natural language processing. pp. 891–896 (2013)
10. Kiros, R., Zhu, Y., Salakhutdinov, R.R., Zemel, R., Urtasun, R., Torralba, A., Fidler, S.: Skip-thought vectors. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015), https://proceedings.neurips.cc/paper_files/paper/2015/file/f442d33fa06832082290ad8544a8da27-Paper.pdf
11. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Xing, E.P., Jebara, T. (eds.) Proceedings of the 31st International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 32, pp. 1188–1196. PMLR, Bejing, China (22–24 Jun 2014), https://proceedings.mlr.press/v32/le14.html
12. Liu, H., Singh, P.: Conceptnet—a practical commonsense reasoning tool-kit. BT technology journal 22(4), 211–226 (2004)

13. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-based and knowledge-based measures of text semantic similarity. In: Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1. p. 775–780. AAAI'06, AAAI Press (2006)
14. Mobile Application Security Alliance: IoT-1001-1 v2.0 Image Monitor System Information Security Standard - Part 1: Information Security Requirements (2021)
15. Mohamed, M., Oussalah, M.: A hybrid approach for paraphrase identification based on knowledge-enriched semantic heuristics. Language Resources and Evaluation 54, 457–485 (2020)
16. OWASP IoT Security Team: OWASP Internet of Things (IoT) Top 10 2014. (2014)
17. OWASP IoT Security Team: OWASP Internet of Things (IoT) Top 10 2018. (2018)
18. OWASP IoT Security Team: OWASP Top 10 vulnerability 2021. (2021)
19. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. CoRR abs/1802.05365 (2018), http://arxiv.org/abs/1802.05365
20. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 3982–3992. Association for Computational Linguistics, Hong Kong, China (Nov 2019), https://aclanthology.org/D19-1410
21. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
22. Swamy, S.N., Kota, S.R.: An empirical study on system level aspects of internet of things (iot). IEEE Access 8, 188082–188134 (2020)
23. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th annual meeting of the association for computational linguistics. pp. 384–394 (2010)
24. Wang, Z., Mi, H., Ittycheriah, A.: Sentence similarity learning by lexical decomposition and composition. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 1340–1349. The COLING 2016 Organizing Committee, Osaka, Japan (Dec 2016), https://aclanthology.org/C16-1127

Yu-Chih Wei is an Associate Professor in the Department of Information and Finance Management at the National Taipei University of Technology. He holds a Ph.D. in Information Management from National Central University, and a B.S. and a M.S. in Information Management from YuanZe University. His research interests include FinTech security, health informatics security, ISRA, SupTech, VANET security, information security management, and business continuity management. Before pursuing an academic career, Dr. Wei was a researcher at the Information & Communication Security Laboratory of Chunghwa Telecom Co., Ltd.

Yu-Chun Chang received his M.S. degree in Department of Information and Finance Management, National Taipei University of Technology in 2023. His research interests include information security and text mining.

Wei-Chen Wu is Assistant Professor in the Department of Finance at the National Taipei University of Business. He received his Ph.D. degree in Information Management from National Central University in 2016. From 2020-2021, He was Assistant Professor in the Department of Finance at the Feng Chia University. From 2008-2016, he was also

Assistant Professor and Director of the Computer Center at Hsin Sheng College of Medical Care and Management. His teaching interests lie in the area of programming languages, ranging from theory to design to implementation and his current research interests include blockchain technology, fintech cybersecurity, network security, and deep learning. Wei-Chen Wu has collaborated actively with researchers in several other disciplines of computer science. He has served on many conference and workshop program committees and served as the workshop chair for Frontier Computing Conference (FC2017 FC2021) and Machine Learning on FinTech, Security and Privacy Conference (MLFSP2019 MLFSP2023).

Received: August 22, 2023; Accepted: November 21, 2023.