Extracting Co-occurrences of Emojis and Words as Important Features for Human Trafficking Detection Models

Chawit Wiriyakun and Werasak Kurutach

Abstract—**Human trafficking is an illegal activity and a major problem of humanity that governments of most countries are trying to prevent. Recently, traffickers have been using social media on the Internet to promote and advertise their business, especially prostitution. Emoji as well as some special words, semantically recognized only in their community, have been used to conveying messages in their advertising communication. This makes it harder for law enforcement officers to track and prevent the activities. In this paper, we propose a feature selection approach focusing on the co-occurrence of emoji and important words for training machine learning (ML) models to detect human trafficking advertisement on social media. In our experimentation, we employed 3 ML models in order to compare our work against the baseline models of E. Tong *et al.* using the trafficking-10k data set. The result has shown that our method significantly outperforms the other's in terms of the F1-score.**

Index Terms— **Human Trafficking, Emoji, Machine Learning.**

I. INTRODUCTION

HUMAN trafficking is an important long-standing problem. In the past, procurers communicated with their customers in certain physical locations. Therefore, it is not too difficult for law enforcement to track down their illegal business. However, in this era, they use social media on the Internet for advertising their businesses, so it is easier and more secure to communicate to their prospect customers. In addition, they can hide advertising information by using specific words and emoji so that law enforcement officers cannot detect and track them easily. During the past few years, many researchers have investigated how to employ machine learning algorithms to automatically detect prostitution advertisements on social media [1-3]. Unfortunately, most of them ignored emojis and concentrated on the text formats only. In those cases, they removed all emojis from the training data sets before selecting important features to train models, However, in the literature of psychology, it discovered that the use of emojis in human communication via social media can influence people's moods and understandings [4]. Prostitution advertisements via social media also use emojis in their messages to promote their illegal businesses and to hide obvious words from being detected by officials. In order to improve the accuracy of automatic detection by ML algorithms, we believe that co-occurrences of emojis and certain words are also important features to consider.

The rest of the paper is organized as follows. Related works are discussed in Section II. Section III briefly describes the data set used in our experimentation. Section IV proposes our method. Section V describes the experimentation and its result. Section VI discusses the experimental result. Finally, the conclusion and future work are presented in Section VII.

Chawit Wiriyakun, Faculty of Engineering and Technology, Mahanakorn University of Technology, Bangkok, 10530, Thailand, e-mail: chawitmild@gmail.com.

Werasak Kurutach, Faculty of Engineering and Technology, Mahanakorn University of Technology, Bangkok, 10530, Thailand, e-mail: werasak@mut.ac.th.

II. RELATED WORK

The research area related to detecting online human trafficking activities has been received more attentions from researchers. R. McAlister [5] proposed a web scraping technique to collect human trafficking data in the Romanian language. This approach can collect data of Romanian advertisements which are more likely related to human trafficking. They translated key phrases that are frequently contained in erotic job advertisements in English into Romanian. Then, the Romanian key phrases were used for searching Romanian advertisements that possibly relate to human trafficking activities. Finally, they used the web scraping technique to collect data. However, the approach has problems in data reliability because they do not have experts to validate data. S. Roshan *et al.* [6] presented an approach to collect human trafficking data via mobile applications. Users send the data from the mobile applications to a server while they are in an anonymous state. After that, researchers used the data to research human trafficking detection. However, the approach still has problems in duplicate and reliability of data. M. Hultgren *et al.* [7] used knowledge management to analyze advertisements which are collected from backpage.com and, then, extracted the keywords and essential data, e.g., the nationalities and the locations of victims. However, the approach only focused on extracting insights from data. The approach does not include a system to assist human trafficking activities and rescuing trafficking victims. R. Kapoor *et al.* [8] presented a framework to extract location data from websites that relate to human trafficking. The framework preprocesses the data, e.g., remove HTML tags. Next, extract geo tags from texts by their semantic lexicons dictionary. After that, analyze the location from geo tags data for human trafficking. However, the approach still has a problem in some cases, e.g., same alternate name of city. A. Mensikova and C. A. Mattmann [9] presented an approach that applied sentiment analysis to analyze human trafficking in escort data collect from webs. In training the model of sentiment analysis, they used a reviews data set. The approach has a problem in reducing the data's bias. M. Hernández-Álvarez [10] presented an approach to detect human trafficking information on Twitter. First, they collect data via Twitter API. Second, they clean the data, e.g., remove unreadable data. Third, they define conditions for feature selection. Finally, they make classification models by the features. So the approach can detect tweets that have possibly relate to human trafficking. However, the approach has a limit in emoticon analysis because tweets have many various emoticons. M. Diaz and A. Panangadan [11] present approach to build data sets for sex trafficking business identification by Natural Language Processing (NLP). They combine the Rubmaps data set that reviews illicit services of businesses with the Yelp data set that is an online reviews data set by mapping address and location. Next, they prepare data, e.g., remove stop words, tokenization, and TF-IDF weighting. Finally, they use the features to create models. The model can identify businesses that have illicit services. However, the approach has a limit in mapping data sets by location name. Because not only some business location address is unclear but also changes locations much time.

Based on our review of related works, most researchers have focused only on data in the text format by excluding emojis which are normally available in many advertisements. This exclusion may affect the performance of models in detecting human trafficking activities. Therefore, we are interested in investigating the inclusion of emojis as an important feature to create a model.

III. DATA SET

This research uses the data set of trafficking-10k provided by Marinus Analytics [1]. The data set contains example advertisements and their levels of classification related to human trafficking activities. The data set was collected from backpage.com that is a classified advertising website. The creation time of advertisements not include in the data set. Originally, the levels of classification in the data set are provided by domain experts. The level definition is shown in Table I.

TABLE I
LEVEL DEFINITION

Level	Definition
0	Strongly Likely Not Trafficking
1	Likely Not Trafficking
2	Weakly Likely Not Trafficking
3	Unsure
4	Weakly Likely Trafficking
5	Likely Trafficking
6	Strongly Likely Trafficking

In the trafficking-10k data set, there are 7 levels (or labels) of data. But, in our work, we are interested in only binary classification, i.e., we need to classify an example data as either "Trafficking" or "Not Trafficking". Therefore, we need to redefine the levels by labeling all example data in levels 0, 1 and 2 as "Not Trafficking" ("0") and in levels 4, 5 and 6 as "Trafficking" ("1"). All data in level 3 are excluded from the data set in order to avoid uncertain data which may be regarded as noise.

IV. APPROACH

In this section, we will introduce a method to extract an important kind of features, co-occurrences of emojis and textual words, that could improve the performance of classification models in detecting human sex trafficking. Our approach consists of 5 main steps: Data Preparation, Emoji to Text Conversion, Tokenization, Co-Occurrence List Creation, and Feature Addition Creation (see Fig. 1). Each step will be explained in each following subsection.

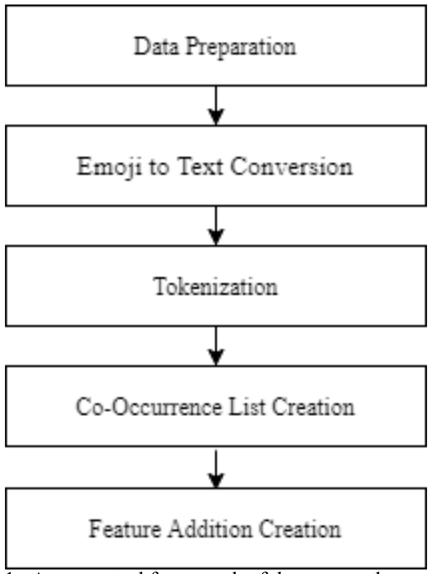

Fig. 1. A conceptual framework of the proposed approach

A. Data Preparation

This step is to prepare all example data in the data set for the feature extracting process. It is composed of five tasks. Firstly, all HTML tags have to be removed from the texts. Secondly, we convert all letters into lowercase ones. Thirdly, we change word formats with the same basic meaning into the same simple form. For instance, "study" and "studies" have the same meaning and, therefore, "studies" is modified to "study". This step is called lemmatization. Fourthly, all numbers are to be replaced by '<number>'. Finally, all stop words, the most common words in English, such as "is", "am", "are", "a", "an" and "the", are removed. Basic techniques in performing these tasks are generally well known and are not necessary to describe in details in this work. A summary of the steps in data preparation is as follows:

Step 1. HTML Removal
Step 2. Convert to Lowercase
Step 3. Lemmatization
Step 4. Replace number by '<number>'
Step 5. Stop words Removal

Some example results of data preparation are shown in Table II.

TABLE II
DATA PREPARATION EXAMPLE

Before	After
Sexy girl for you 🛌	sexy girl you 🛌
👩 Girl wait to serve you	👩 girl wait serve you
<h3> Friend 👤 24 time<h3>	friend 👤 <number> time

B. *Emoji to Text Conversion*

Emojis are popular and often used in advertisements because they can promote customers' positive moods [4]. Moreover, they are better at conveying meaning than only using textual words. Therefore, sex-trade procurers always use both emojis and textual words in their advertisements to promote the business and to hide their activities from law enforcement. In the past, most detection models focused on detecting activities only from the textual words. As a result, sex-trade advertisements can escape from detection. In this work, we regard emojis as a very important component that cannot be ignored. Therefore, rather than removing emoji, we convert them into texts, e.g., 🏩 into :love hotel: and 👧 into :girl:. The corresponding texts of the emojis are defined by Unicode Consortium [12]. Table III shows some example results of this conversion step.

TABLE III
CONVERT EMOJI TO TEXT EXAMPLE

Before	After
sexy girl you 🏩	sexy girl you :love hotel:
👧 girl wait serve you	: girl: girl wait serve you
friend 👩 <number> time	friend :woman: <number> time

C. *Tokenization*

Preliminarily, the advertisements are to be tokenized based on the whitespace. However, some advertisements do not use the whitespace when having emojis in the content, e.g., 👧you🏩. Therefore, it is necessary to add a whitespace between the word and the emoji before tokenizing the content. We can separate this task into four steps. Firstly, we tokenize the content based on the whitespace. Secondly, we check each token to find the one that contains a word and its coalescing textual word of emoji (converted from emoji). Thirdly, we add a whitespace between the words and the textual word of emoji in such a token. Finally, we re-tokenize the content of the token received from the previous step. An example of this task can be seen in Fig. 2.

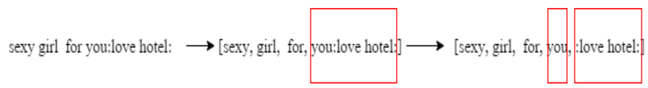

Fig. 2. An example of tokenization

The commas are separators of tokens, so the output in the example has five tokens. Table IV shows some examples of data before and after the process of this task.

TABLE IV
BEFORE AND AFTER TOKENIZATION

Before	After
sexy girl you :love hotel:	[sexy, girl, you, :love hotel:]
:girl: girl wait serve you	[:girl:, girl, wait, serve, you]
friend :woman: <number> time	[friend, :woman:, <number>, time]

Tokenization procedure is illustrated in Algorithm 1. We use the return tokens from Algorithm 1 to create a co-occurrence list in the next task.

```
Algorithm 1: Tokenization
procedure Tokens(advertisement)
  Tokens[] ← TokenizeBySpace(advertisement)      ▷ Tokenize
content by Whitespace
  ReturnTokens[]
  N ← length(Tokens)                              ▷ Tokens length
  for i ← 1 to N do
    if TextEmojiAndWord(Tokens [i]) then
      WordEmojiText ← AddSpaceEmojiText(Tokens [i])) ▷ Add
space between word and text that converted from emoji
      Item[] ← TokenizeBySpace(WordEmojiText)
      O ← length(Item)                      ▷ Tokens Item length
      for j ← 1 to O do
        ReturnTokens.Append(Item[j])
    else
      ReturnTokens.Append(Tokens[i])
  return ReturnTokens                  ▷ For make list in next task
```

D. *Co-Occurrence List Creation*

This task creates a co-occurrence list containing words and texts (converted from emoji). A co-occurrence is a pair of a token which is a text converted from an emoji and its previous or next token, e.g., girl :love hotel:, and :woman: time. This process can be seen in Algorithm 2.

```
Algorithm 2: Co-Occurrence List Creation
procedure ListCreation(tokens)
  Tokens[] ← tokens
  ReturnList[]
  N ← length(Tokens)                              ▷ Tokens length
  for i ← 1 to N do
    if isTextFromEmoji(Tokens[i]) then    ▷Check if text that
converted from emoji in token
      if isNotNull_And_NotUndefined (Tokens[i-1]) then
        if isWord(Tokens[i-1]) then
          ReturnList.Append(Tokens[i-1] + " " +Tokens[i])
      if isNotNull_And_NotUndefined (Tokens[i+1]) then
        if isWord(Tokens[i+1]) then
          ReturnList.Append(Tokens[i] + " " + Tokens[i+1])
  return ReturnList
```

An example of this task is shown in Fig. 3.

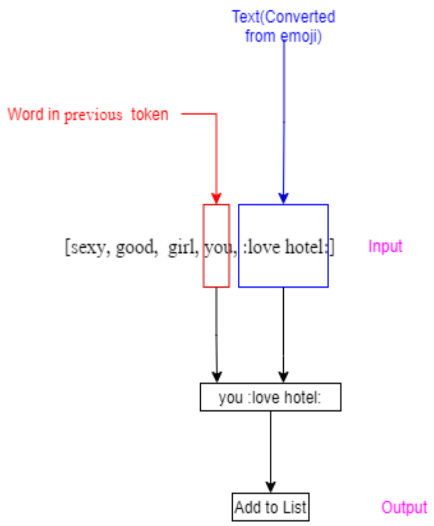

Fig. 3. An example of co-occurrence list creation

E. Feature Addition Creation

In this task, we find the inverse document frequency (IDF) of each co-occurrence in the list. The IDF can be evaluated by Equation (1). If the IDF value of a co-occurrence is less, it means such a co-occurrence is used more often. Therefore, any co-occurrence that is used more often than average will be chosen as an important feature.

$$IDF(A,C) = \log\frac{A}{C} \quad (1)$$

where A is the total number of advertisements that contain all co-occurrences in the list, and C is the total number of advertisements that contain a co-occurrence that we want to calculate its IDF value.

We select all co-occurrences that is used more often than the average to be features in model creation.

V. Experimentation

In our experimentation, the data set was trafficking-10k. We carried out the process of data preparation as described in Subsection *A* of Section IV. Then, we used our approach as explained in Subsections *B* – *E* to extract important features which include co-occurrences of word-emoji pairs for training or creating binary classification models. We choose three kinds of models for our experimentation – Random Forest, Logistic Regression and Linear SVM in order to compare the result with the work of E. Tong *et al.* [1] who owns the data set of trafficking-10k. In the work, they used the technique of Bag of Words for feature extraction to train the three baseline models. In our comparison, we used the same data set and the same values of all learning parameters. Due to the unbalanced data set, the F1-score was employed for measuring and comparing the performances. It can be calculated by Equation (2).

$$F1 = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (2)$$

The experimental result is shown in Table V.

TABLE V
EXPERIMENTAL RESULT

Model	E. Tong et al.'s approach (%)	Our approach (%)
Random Forest	24.5	**63.3**
Logistic Regression	24.5	**64.8**
Linear SVM	24.5	**61.3**

VI. Discussion

Based on the result shown in Table V, it can be seen that introducing emoji-word co-occurrences as additional features for training classification models can improve their performances significantly. The performances are measured in terms of F1-score which is normally used in the field of machine learning. In the table, the result of our approach is compared to E. Tong *et al.*'s [1]. When using Random Forest model, F1-Score of our approach is 63.30% whereas the E. Tong et al.'s is 24.50%. Obviously, the performance of the model has improved tremendously. Moreover, when experimenting with Linear Regression and Linear SVM models, the improvement scale is similar to Random Forest. The overall result seems to prove our contention that emoji-word co-occurrences are important features for creating classification models to detect human trafficking advertisements in prostitution.

VII. Conclusion

In this paper, we propose techniques of extracting co-occurrences of words and emojis in human sex trafficking advertisements. We contend that these co-occurrences are an important feature that can improve the performances of machine learning algorithms in detecting online advertisements relating to human trafficking, especially prostitution. Furthermore, we experiment our co-occurrence feature extraction with three ML algorithms, Random Forest, Logistic Regression and Linear SVM, using the trafficking-10k data set. The experimental result has shown that the performances of the three algorithms have improved significantly compared to E. Tong *et. al.*'s work [1].

In the future, we will investigate other kinds of co-occurrences such as word-word co-occurrences that may be important for classifier induction.

Acknowledgment

We would like to thank Cara Jones for providing us the human trafficking data set.

References

[1] E. Tong, A. Zadeh, C. Jones, and M. Louis-Philippe, "Combating Human Trafficking with Multimodal Deep Models," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017.pp. 1547-1556.

[2] J. Zhu, L. Li, and C. Jones, "Identification and Detection of Human Trafficking Using Language Models," in: *European Intelligence and Security Informatics Conference (EISIC)*, 2019. pp. 24-31.

[3] M.C. Lee, C. Vajiac, A. Kulshrestha, S. Levy, N. Park, C. Jones, C., R. Rabbany, and C. Faloutsos, "InfoShield: Generalizable Information-Theoretic Human-Trafficking Detection," in *37th IEEE International Conference on Data Engineering (ICDE)*, 2021.pp. 1116-1127.

[4] Q. Bai, Q. Dan, Z. Mu, and M. Yang, "A Systematic Review of Emoji: Current Research and Future Perspectives," in *Frontiers in Psychology (10)*, 2019.

[5] R. McAlister, "Webscraping as an Investigation Tool to Identify Potential Human Trafficking Operations in Romania," in *Proceedings of the ACM Web Science Conference (WebSci '15)*, 2015. pp. 1–2.

[6] S. Roshan, S. V. Kumar, and M. Kumar, "Project spear: Reporting human trafficking using crowdsourcing," in *4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, 2017. pp. 295-299.

[7] M. Hultgren, M. E. Jennex, J. Persano, and C. Ornatowski, "Using Knowledge Management to Assist in Identifying Human Sex Trafficking," in *49th Hawaii International Conference on System Sciences (HICSS)*, 2016. pp. 4344-4353.

[8] R. Kapoor R., M. Kejriwal, and P. Szekely, "Using Contexts and Constraints for Improved Geotagging of Human Trafficking Webpages," in *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, Article 3. 2017.

[9] A. Mensikova, and C.A. Mattmann, "Ensemble Sentiment Analysis to Identify Human Trafficking in Web Data," in *Proceedings of ACM Workshop on Graph Techniques for Adversarial Activity Analytics*, 2018. 6 pages.

[10] M. Hernández-Álvarez, "Detection of Possible Human Trafficking in Twitter," in *International Conference on Information Systems and Software Technologies (ICI2ST)*, 2019. pp. 187-191.

[11] M. Diaz and A. Panangadan. "Natural Language-based Integration of Online Review Data sets for Identification of Sex Trafficking Businesses," in *IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, 2020.pp.259-264.

[12] Full Emoji List, https://unicode.org/emoji/charts/full-emoji-list.html, last accessed 2021/08/04.

Chawit Wiriyakun received the B.Eng. (Information and Communication Engineering) degree from Mahanakorn University of Technology in 2011. He received the M.S. (Information Technology) degree from Kasetsart University in 2014. He is currently an engineer at the National Science and Technology Development Agency (NSTDA) in Thailand. He is also a Ph.D. Student at the Faculty of Engineering and Technology, Mahanakorn University of Technology.

Werasak Kurutach received his PhD in Computer Science and Engineering from The University of New South Wales, Australia, in 1996. Currently, he is an associate professor in computer engineering at Mahanakorn University of Technology, Bangkok, Thailand.