Identifying Sources of Human Trafficking Within Online Escort Advertisements Written in Spanish

A thesis submitted for the degree of Master in Industrial Engineering
Universidad de Los Andes, School of Engineering

Paula Rodríguez Díaz

June, 2021

Abstract

The problem of Entity Resolution from online escort advertisements has been well studied using Information Extraction (IE) frameworks with English based regular expressions to isolate sources of human trafficking. However, these approximations create noisy features, they cannot be directly used on escort advertisements written in other languages, and require high computational capacity to perform pairwise comparisons. This thesis proposes an entity resolution pipeline based on a Contrastive Learning (CL) and clustering framework that is able to identify possible sources of human trafficking by extracting clusters from text embeddings. The proposed pipeline is tested on escort advertisements written in Spanish obtaining an AUC of 0.93 for the CL framework, from which we are able to identify entities with domain specific characteristics.

1 Introduction

Human trafficking, commonly found within organized criminal networks of prostitution, is a challenging law enforcement problem and a worldwide concern that keeps growing while finding its way amid multiple digital platforms. Forms of exploitation among human trafficking victims include sexual exploitation, forced labor, criminal activities, forced marriages, amongst others. However, most human trafficking victims are trafficked for the purpose of sexual exploitation, representing 50% of human trafficking cases detected globally in 2018 and up to 64% in South America according to UNODC (2020).

As the world continues to transform digitally, traffickers are taking advantage of online platforms to perpetuate human trafficking crimes by using these technologies to recruit, advertise, and exploit victims. The way law enforcement can use these digital traces can be divided and understood from two different action paths: prevention and identification. While recruitment on social media and webpages may assist law enforcement in preventing future cases of human trafficking, advertisement and online exploitation can lead to identifying current trafficking cases.

As stated in the Global Report on Trafficking in Persons by UNODC (2020), "traffickers increasingly use internet technologies to advertise the services resulting from their victims' exploitation. Examples of advertisements used to exploit victims include those on classified listing sites, such as *Backpage* and similar, or on social media platforms and applications". Although the classified listing website, www.backpage.com, was seized by the US Federal Bureau of Investigation in April 2018 [The United States Department of Justice (2018)], there has not yet been any legal action taken against similar listing sites operating in other countries such as *Mileroticos* –which by itself operates in Mexico, Colombia, Chile, Brazil, Italy, and Spain.

Given the current scenario of online publicly available data that possibly traces human trafficking activities, researchers such as Konrad et al. (2016) and Caulkins et al. (2019) have made a call to the operations research and analytics community to address the growing issue of human trafficking from data-driven solutions. The identification of human-trafficking patterns on online platforms using network analysis, natural language processing, and standard machine learning classification algorithms has been well studied for online information written in English and, particularly, from the United States [Dubrawski et al. (2015), Nagpal et al. (2015), Rabbany et al. (2018), Kejriwal and Szekely (2017), Kejriwal et al. (2017), Alvari et al. (2017), Ibanez and Suthers (2014)]. However, given that most of these models are based and trained on English regular expressions for fulfilling information extraction tasks, frameworks for identifying human trafficking patterns such as RedThread by Rabbany et al. (2018) or FlagIt by Kejriwal et al. (2017) will not necessarily work for identifying trafficking victims in non English speaking countries.

State of the art models that identify human trafficking cases in English online data can be divided into two groups by their final task: (1) models that identify individual victims and (2) models that identify organized trafficking groups. Regarding the first group, Dubrawski et al. (2015) in its section *Highlighting Data of Interest* trained a Random Forest Classifier on specific features of online advertisements that contained phone numbers against known traffickers' phone numbers, and Kejriwal et al. (2017) combined lightweight expert system and semi-supervision with unsupervised text embeddings to create indicators that are highly correlated with human-trafficking by using red flags such as movement between cities and advertisement of multiple women.

Models that identify organized trafficking groups within online advertisements include Dubrawski et al. (2015) in its section on *Identifying and Tracking Entities* where they use entity-resolution algorithms on weak features obtained by English based information extraction tasks that are able to trail a sequence of highly textually homogeneous advertisements. Nagpal et al. (2015) uses a regular expression information extractor constructed via GATE and a classification algorithm to predict if two advertisements are from the same source (same phone), and Rabbany et al. (2018) developed RedThread, an efficient solution for inferring related and relevant nodes while incorporating the user's feedback to guide the inference on case building for combating human trafficking.

To the best of the author knowledge, the only data analysis methods used for identifying human trafficking patterns in online data written in Spanish have been initially studied by Alvarez and Burbano Acuña (2017), which from 2017 until today have remained in developing information extraction techniques to create a corpus of Spanish texts related to human trafficking. Thus, no further developments have been made in terms of identifying patterns related traffickers or victims within online data in Spanish.

In Colombia, which will be the focus of this work, the judiciary and law enforcement officials will most probably work toward the judicialization of organized trafficking groups than of isolated cases of trafficked victims. As such, the former Deputy Prosecutor for Children and Adolescents from the General Attorney's Office of Colombia[1], in an interview with the author, stated that: "when there's a situation of a woman who consents it (a sexual service) because she is being paid, in general, there is no interest from anyone. What's left then? When it is contrary to the will, when it has been deceived, when it has been taken to that place under deceit and is subjected in its will to remain there (...) where in addition the pimp keeps the money. In that case, there is a lot of interest from the judiciary and law enforcement officials." Furthermore, a wide array of criminals and groups are involved in trafficking persons and, from this, more structured organizations may traffic more victims for longer periods [UNODC (2020)].

In this paper, we explore the use of a semi-supervised text similarity approach with a non predefined language to identify sources of advertisement which, as stated before, can potentially be related to organized criminal groups involved in human trafficking. The proposed scheme is implemented on advertisements written in Spanish obtained during 90 consecutive days from the listing site www.mileroticos.com in the location of Colombia. Although the proposed approach is tested on Spanish texts, it can be extended with no further methodological changes to any other language since it does not depend on language-specific regular expressions for feature extraction.

[1]Ex-Fiscal delegado para infancia y adolescencia (2016-2020)

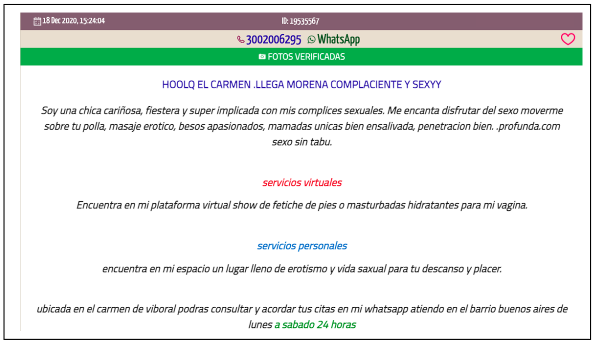
This project contributes to the literature on machine learning based models with applications to human trafficking by proposing an entity recognition scheme for online escort advertisements that does not depend on a predefined language nor an information extraction scheme which usually adds noise to the data extracted [Kejriwal and Szekely (2017)]. The application and implementation presented in this paper also addresses one of the most relevant tasks for the judiciary and law enforcement officials combating human trafficking on online platforms in Colombia as it identifies possible trafficking networks rather than isolated cases.

2 Data Description

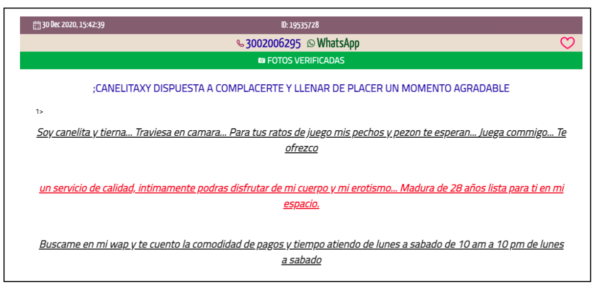
The escort advertisement download from www.mileroticos.com was done by daily scraping during 90 consecutive days beginning March 13, 2021. This technique consists of downloading the HTML code from the website pages and processing it for information extraction. The download and information extraction process code was developed using Python. The web scraping was done by creating a program that goes to the page of interest, extracts all the information from the advertisements and turns to the next page. Information extraction is done by traversing the HTML code with XPATH. To automate the execution of the program, we configured it to run automatically every day in a cloud service.

From March 13, 2021 to June 13, 2021 (90 days) we scrapped a total of 260,493 unique advertisements. The uniqueness was defined by the url associated to each advertisement and not by it's content. From the advertisements downloaded, we noticed that most of them remain on the webpage for 1 to 5 days, making it necessary to create a daily scrapping in order to have complete information.

Figure 1 shows two escort advertisements posted on the scraped website that have the same phone number associated but different urls and text content. The implemented web scraping extracts the title, text description (body), phone number, location, and listing date associated to each publication. However, we will only use the phone number and text description (body) for the proposed Entity Resolution (ER) approach.



(a) Escort advertisement posted on December 18, 2020.



(b) Escort advertisement posted on December 30, 2020.

Figure 1: Escort advertisements posted on www.mileroticos.com on different days of December 2020 with the same phone number but with different text content.

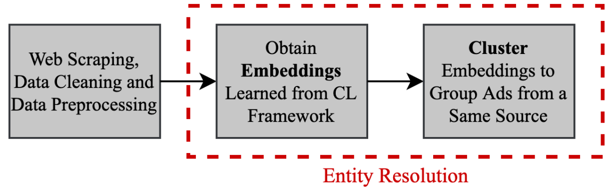


Figure 2: The proposed pipeline for Entity Resolution of online escort advertisements.

3 Entity Resolution

3.1 Definition

Entity Resolution (ER) is the task of identifying and matching records (also known as instances) that come from a same entity. State of the art models for isolating sources of human trafficking from online escort advertisements such as the one proposed by Nagpal et al. (2015) use (i) Information Extraction (IE) frameworks with English based regular expressions that cannot be directly applied to escort advertisements in other languages such as Spanish [Alvarez and Burbano Acuña (2017)], (ii) use text features extracted from IE frameworks based on regular expression which are time consuming, they require a rigorous validation scheme [Dubrawski et al. (2015)], and create noisy features [Kejriwal and Szekely (2017)], and (iii) exhaustive pairwise comparisons between advertisements which require high computational capacity.

The ER framework proposed in this paper, which is outlined in Figure 2 and described in detail in the following sections, addresses challenges (i) and (ii) by using a Siamese BLSTM network trained with a Contrastive Learning framework to learn advertisements' embeddings that capture both specific features and underlying characteristics from the texts without using a pre-defined language IE framework. The following use of clustering methods on the learned embedded space allow to group advertisements from a same source without requiring pairwise comparisons (iii).

The ER problem can be formally defined as follows. Consider the set of vertices $\mathscr{V} = \{v_1, v_2, \ldots, v_n\}$ which is the set of escort advertisements (defined by their text content and phone number) and $\mathscr{E} =$

$\{(v_i, v_j), (v_j, v_k), ..., (v_k, v_l)\}$ the set of edges between advertisements that come from a same entity (same phone number). For the ER problem we want to train a learning function M such that $M(v_i, v_j) = 1$ if $(v_i, v_j) \in \mathscr{E}$ and $M(v_i, v_j) = 0$ if $(v_i, v_j) \notin \mathscr{E}$.

Under the assumptions that:

1. advertisements with the same phone number come from the same entity,
2. phone numbers are related to at most one entity and entities may use more than one phone number, and
3. entities have writing styles that, when being strong enough, can uniquely determine the correspondence between instances and entities,

we propose a function M that is trained using a contrastive learning framework along with clustering methods while assuming that $M(v_i, v_j) = 1$ for all pair of advertisements v_i, v_j that have the same phone number. By evaluating the learned function M in our complete dataset, we are able to determine if two advertisements v_i, v_j with different phone numbers are or not connected i.e. come or not from the same entity.

Our proposed method for learning the function M uses the texts body descriptions and phone numbers from escort advertisements in a Contrastive Learning + Clustering framework that does not depend on language based regular expressions nor exhaustive pairwise comparisons. We assume as ground truth that advertisements posted by a same phone number come from the same source. With our Entity Resolution pipeline we are interested in finding groups of advertisements that potentially come from the same entity but use different phone numbers.

3.2 Advertisements Embeddings

3.2.1 Contrastive Learning Framework

The proposed Contrastive Learning Framework trains a Bidirectional LSTM (BLSTM) layer to obtain text embeddings that maintain a defined distance between advertisements according to their source i.e. vector representations of advertisements that come from a same source are closer in the embedded space than advertisements coming from a different source. For our specific application, same source advertisements correspond to advertisements written by a same author or organization, which can potentially represent organized trafficking groups.

Contrastive Learning frameworks such as the ones proposed by Neculoiu et al. (2016) and Chen et al. (2020) provide a self-supervised way of obtaining sentence embeddings that reflect entities' underlying characteristics for our specific application and do not require initially labeled data. Unlike word embeddings, the highest performing solutions for encoding the underlying meaning expressed in a sentence require labelled data [Giorgi et al. (2021)]. The self-supervised training framework allows us to obtain high performing sentence embeddings without initially having labelled data but by labelling it in a contrastive manner.

The proposed text embedding model with a contrastive learning framework based on Neculoiu et al. (2016) and outlined in Figure 3, consists of a siamese recurrent network which is defined as follows:

- The first layer of the network receives two inputs $(x^a_1, ..., x^a_{N_a})$ and $(x^b_1, ..., x^b_{N_b})$, which correspond to initial word embeddings of the advertisements a and b. In our case, we trained two models: one with a given word embedding (we used Word2Vec developed by Mikolov et al. (2013)) and one with an initial trainable embedding layer.

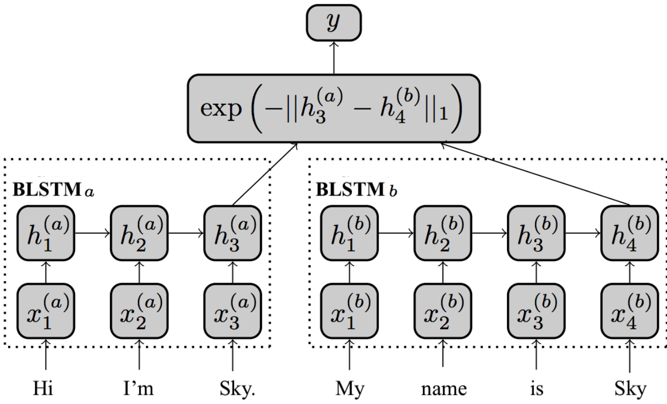


Figure 3: The siamese BLSTM structure consists of three layers: a right and left input or embedding layer, a right and left BLSTM layer, and a final joining layer for the contrastive function. Image adapted from Neculoiu et al. (2016).

- The second layer is a Bidirectional LSTM (BLSTM). This layer is specialized in exploiting the sequential structure of text a, as well as text b. Its output is a vector representation (in 128 dimensions) of the text, which will be denoted as h^a_{Na} and h^b_{Nb}, respectively. If the $BLSTM$ layer is considered as a function, then it would be applied as follows:

$$BLSTM(a) = BLSTM((x^a_1, \ldots, x^a_{N_a})) \quad (1)$$
$$= h^a_{Na} \quad (2)$$

  where $h^a_{Na} \in \mathbb{R}^{128}$.

- The last layer applies the contrast function. This contrast function returns a number between 0 and 1. In case of being 1, the texts come from the same entity, in case of being 0 the texts come from different entities. For this case the contrastive function used is:

$$g(h^a_{Na}, h^b_{Nb}) = \exp(-||h^a_{Na} - h^b_{Nb}||) \quad (3)$$

After training the model on advertisement pairs with the same phone number (postive pairs) and sufficiently different advertisements assumed to come from different entities (negative pairs), we are able to obtain text embeddings of 128 dimensions by using the intermediate $BLSTM$ layer.

3.2.2 Positive and Negative Pairs Sampling Scheme

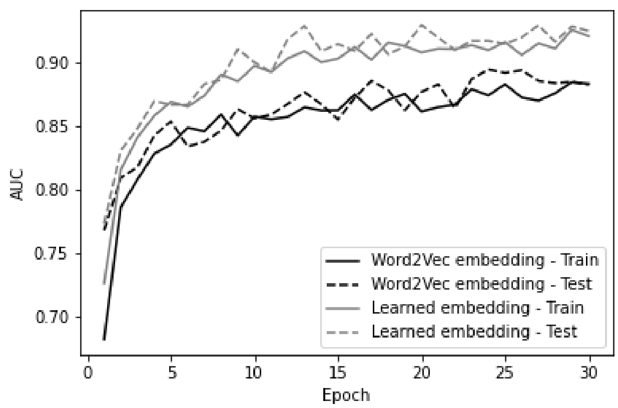
To train the siamese network structure defined in the previous section we label pairs of texts in a semi-supervised manner. We create pairs of texts coming from a same entity (having the same phone number) and mark them as $y = 1$. Likewise, we automatedly create pairs of texts that are sufficiently different to assume that they come from different entities which are labeled as $y = 0$.

We use a Term frequency - Inverse document frequency (TF-IDF) matrix to determine which pair of advertisements come from different entities by assuming that those ones with no words in common come from different sources. By pairwise calculating the cosine similarity between the TF-IDF rows we are able to identify the pairs of texts with no common words. We randomly sample these pairs to the same amount of positive pairs in order to have a balanced dataset used for the contrastive learning classification task.

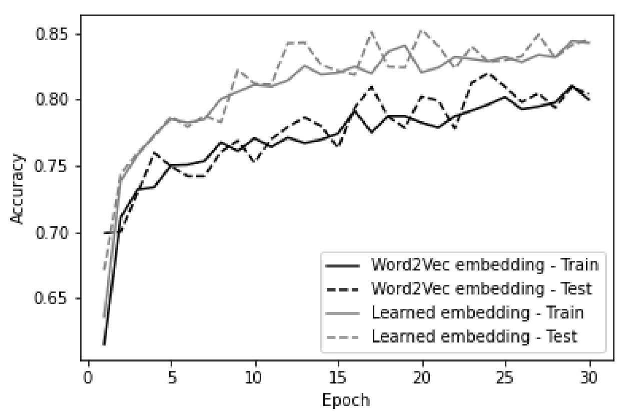
3.2.3 Training and Preliminary Results

We train the siamese BLSTM network with a contrastive learning framework using 164,946 pairs of texts, from which 50% of them correspond to positive pairs i.e. texts from escort advertisements with the same phone number. We use a random 70-30 sample split for the train and test sets.

As mentioned on section 3.2.1, we train two different models that differ in the input received. The first one was trained with an input of pretrained Word2Vec embeddings of dimension 300 and the second one received the texts as input and trains an initial embedding layer as part of the siamese BLSTM network. Figure 4 shows that the model with trainable embedding layer outperformed the model with Word2Vec inputs having a maximum value of 0.930 for the validation AUC and 0.853 for the validation accuracy on epoch 20.



(a) AUC



(b) Accuracy

Figure 4: Performance metrics (AUC and accuracy) for the two versions of the siamese BLST model during each epoch. Darker lines correspond to the model that received Word2Vec embeddings as inputs and lighter lines correspond to the model that has a trainable initial embedding layer.

3.3 Advertisements Clustering

3.3.1 Description

Our escort advertisements dataset consists of 260,493 texts from which 176,458 are unique from their text content and were posted between March 17, 2021 and June 13, 2021. Pairwise comparisons and distance calculations of the text embeddings across the dataset make this problem computationally intractable, specially when doing a constant verification of advertisements. Given that the text embeddings were constructed in a way that instances from a same entity are close in the embedded space, we can expect to isolate instances i.e. learn a function M described in section 3.1, by determining clusters over the embedded space rather than making pairwise distance calculations.

We use Density-based spatial clustering of applications with noise (DBSCAN) [Ram et al. (2010)] with different values ranging from 0.001 to 0.15 for the maximum distance between two samples for one to be considered as in the neighborhood of the other and a fixed value of 12 for the minimum number of samples in a neighborhood. This fixed value was chosen from the fact that advertisements from one same phone number in our dataset were posted every 11.64 days in average which, rounding up to 12, corresponds to posting an advertisement every week on average.

The distribution of euclidean distances for positive pairs (coming from a same phone number) shown in Figure 5 is more concentrated to the left (lower distances) when comparing it to the distribution of distances for negative pairs. Hence, when using a clustering method with euclidean distance in the Siamese + BLSTM embedded space, we can expect to have a high true positive rate when grouping advertisements from a same phone number. Although using a clustering method in this case may result in false positives given the left skew-

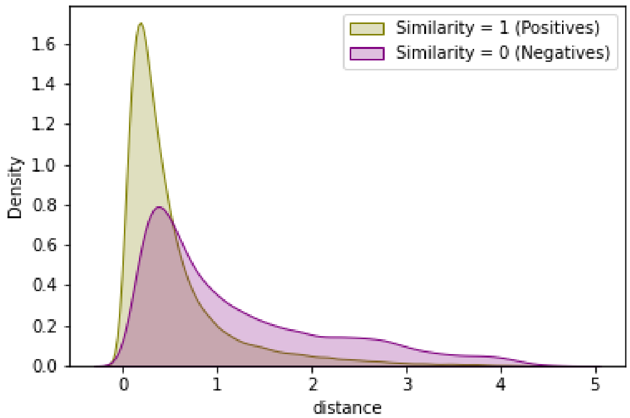


Figure 5: KDE of the euclidean distances of the positive and negative pairs in the Siamese + BLSTM embedded space used to train and test that same network. This pair dataset is the same as described in section 3.2.2.

ness of the negative pairs, we can expect it to be less than the true positives, which are the priority for applications on detecting human trafficking.

We do not expect all advertisements to be part of a cluster containing more than one phone number since there can be individual escorts who use this listing platform by their own will. Hence, algorithms like DBSCAN which identify noise observations, are well suited for solving the entity resolution task in the context of escort advertisement since it is able to isolate individual entities by classifying them as noise.

3.3.2 Tuning parameters

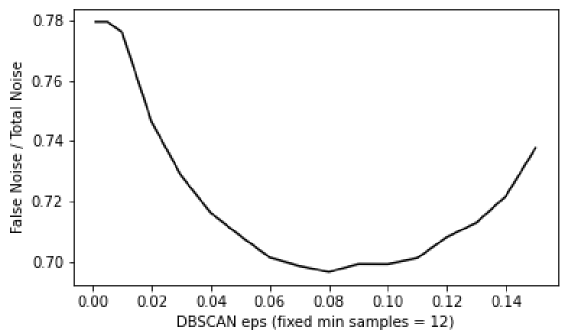
The lack of ground truth of entities and isolated advertisements make it challenging to tune the clustering parameters as there is no true scenario to compare with the clusters generated in the embedded space. However, following our assumption 1 stated in section 3.1 (advertisements with the same phone number come from the same entity), we would at least want to cluster the advertisements coming from repeated phone numbers and not have them labeled as noise.

Figure 6 shows (a) the fraction of advertisements marked as noise that come from repeated phone numbers (some entity) and (b) the fraction of advertisements mistakenly labeled as noise out of the advertisements known to belong to some entity vs eps. We choose 0.8 as the maximum distance between two clustered samples given that advertisements coming from repeated phone numbers should belong to a cluster while isolated advertisements, which are unknown, should be labeled as noise.

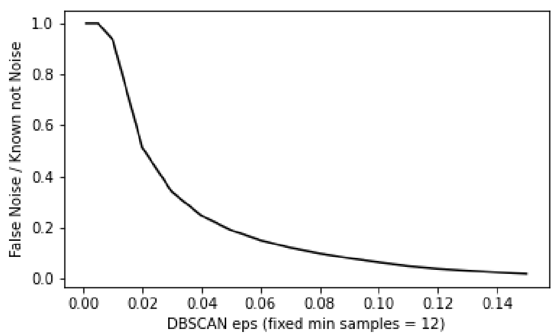
As shown in Figure 6, for the chosen maximum distance between clustered pairs (DBSCAN eps) we have that only 9.91% of advertisements coming from repeated phone numbers are labeled as noise. These observations correspond to the 69.7% of the observations labeled as noise, which es the minimum value obtained for the evaluated maximum distances. For these parameters (eps = 0.8, minimum samples = 12) we obtain 35 clusters and 19,556 advertisements marked as noise. Although the fraction of noise advertisements with repeated phone numbers is high, the distribution for the amount of advertisements for each phone number is more concentrated to the left compared to the whole sample of repeated numbers as shown in Figure 7. This shows that the clustering method is able to group advertisements that come from more recurrent phone numbers better than those that come from phones posting more sporadically.

3.4 Examples of Identified Entities

The embeddings learned from the Siamese BLSTM framework may lack interpretability, specially when being compared to frameworks that extract specific features using regular expressions. The BLSTM layer trained in our contrastive learning framework can learn embeddings that identify specific features or frac-



(a) Fraction of noise mistakenly marked as noise.



(b) Fraction of advertisements coming from some entity mistakenly marked as noise.

Figure 6: Calibration metrics involving positive pairs of advertisements and noise for choosing the maximum distance between two clustered samples (eps) for a fixed amount of 12 minimum samples in a neighborhood.

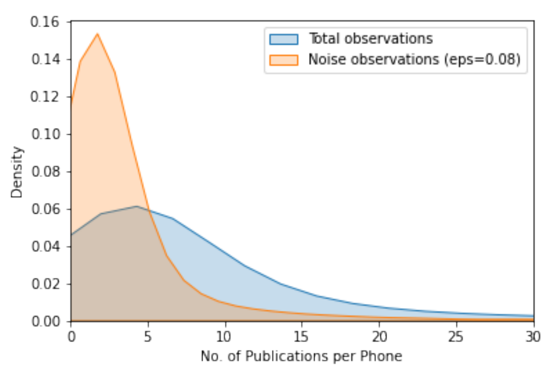


Figure 7: Histogram of number of publications for each repeated number in the whole dataset vs in the noise set.

tions of sentences in the escort advertisements as it can also identify underlying characteristics of the texts that are less interpretable.

Figure 8 shows three advertisements that were assigned to the same cluster using our framework and were listed with different phone numbers. These advertisements belong to a cluster of 28 samples with 4 different phone numbers and show extremely similar textual content, specially at the end of the text, making it seem like a template used with some variations.

Figure 9 presents another example of posts belonging to a same group and having different phone numbers. In this case, although the exact words are not the same and while even the longest common subsequence (`y mis videos. Tienen un costo`) is short and common within multiple advertisements, our framework is able to detect a common structure consisting on starting with mentioning the location, stating the name and the age, and then offering a detailed package with the corresponding pricing and specific venues for the payment transfer.

4 Conclusion and Discussion

This thesis work proposes the use of a semi-supervised text similarity approach with a non predefined language to identify sources of human trafficking within online escort advertisements. The proposed pipeline consists of learning text embeddings that maintain advertisements from a same source close in the embedding space by using a contrastive learning framework that consists of a siamese BLSTM network. The pipeline was tested on online escort advertisements from www.mileroticos.com written in Spanish and located in Colombia. The resulting entity resolution learned function was able to identify clustered advertisements with different phone

```
Hola mis amores soy lorena nueva en palmira
dispuesta para soy muy complaciente descomplicada en
mi encontraras una buena compañia y si quieres
relajarte tengo un rico masaje erótico piel a piel
donde te masajeare toda tu espalda con mis ricos
senos t provaras mi rica vagina y si te gusta
disfrutar el doble dos chicas por 80 mil doble oral
doble vaginal te encantara... mis servicios son
oral vaginal cuatro manos prostatico anal trios
atiendo parejas
```

```
Hola mis amores soy sara dispuesta para ti para que
la pasemos rico soy descomplica cumplo fetiches te
trato como novio ven y prueva mi rico cuerpo lleno
de sensualidad aremos las poces que mas te gusta y
si te gusta disfrutar el doble dos chicas por 80 mil
doble oral doble vaginal te encantara... mis
servicios son oral vaginal cuatro manos prostatico
anal tríos atiendo parejas
```

```
Hola mis amores soy roxana una morena muy fogosa me
encanta el sexo duro para poder disfrutarlo los
mejores movimientos de cadera los tendrás con esta
morenita de rico cuerpo no te vas a arrepentir y
quieras repetir esta experiencia tan rica y la promo
dos chicas por 80 mil doble oral doble vaginal te
encantara bebe... mis servicios son oral vaginal
cuatro manos prostatico anal tríos atiendo parejas
```

Figure 8: Representative results of advertisements from a same cluster and having different phone numbers. This cluster identifies 28 posts with 4 different phone numbers.

```
Popayan y todo el cauca, me dicen *******, tengo 18
años, hoy. Estoy vendiendo mi rico contenido, que
contiene mis fotos y mis videos. Tienen un costo de
1 pack con fotos y videos por 5 mil 3 pack por 10
mil 4 pack por 13 mil Pagos por nequi, davi plata,
bancolombia Escribeme lindo
```

```
Sincelejo, me llamo mariana, tengo 18 años, hoy te
ofrezco mi rico contenido que consta de mis fotoa y
mis videos. Tienen un costo muy economico desde
5000 pesos Los pagos son por medio de Nequi Davi
plata Bancolombia Si estas realmente interesado en
mi contenido casero escribeme
```

```
Barranquilla y todo el Atlántico, me llamo carolina,
tengo 18 años, obtén mi contenido, que contiene mis
fotos y mis videos. Tienen un costo súper económico
de. 1 paquete con fotos y mis videos por 5 mil 3
paquetes por 10 mil 4 paquetes por 13 mil. Pagos
por medio de nequi, Bancolombia, Daviplata.
```

Figure 9: Representative results of advertisements from a same cluster and having different phone numbers. This cluster identifies 12 posts with 4 different phone numbers which are associated to a total of 286 advertisements.

numbers and strong domain specific characteristics without depending on language based methods and using a clutering based approach that does not require to make pairwise comparisons.

The particular task of identifying organized trafficking networks withing escort advertisements rather than individual trafficking cases is of major importance for the judiciary and law enforcement officials from Colombia. The implementation done in this paper shows that the proposed entity resolution pipeline can assist these officials in combating human trafficking by identifying its sources within online escort advertisements. This identification was done under the following assumptions:

1. Escort advertisements with the same phone number come from the same entity,
2. phone numbers are related to at most one entity and entities may use more than one phone number, and
3. entities have writing styles that, when being strong enough, can uniquely determine the correspondence between instances and entities.

Hence, the proposed entity resolution pipeline is limited to identifying entities that have a recurrent writing style or that repeat common information among their different posts. The approach proposed in this paper would not be able to group advertisements from a same entity when the advertisements do not have common information or similar writing styles.

Some of the limitations of our proposed and implemented approach come from not having ground truth labels for advertisements belonging to one same entity and for entities that are directly linked to human trafficking. For the later, we use phone numbers with repeated advertisements as a subset of what would be the ground truth entities. We also create a proxy subset of advertisements that do not belong to a same entity by assuming that advertisements with no words in common necessarily come from different sources. However, we do not have labels that relate escort advertisements with actual human trafficking cases. Hence, although the identified entities can assist the law enforcement on further human trafficking investigations, they may not necessarily be linked to organized trafficking groups as they can be related to other organized crimes –such as scam or fraud– that use listing websites platforms to operate.

One of the biggest challenges when implementing the proposed pipeline is choosing the amount of clusters or, in our specific case, the maximum distance between clustered samples. The lack of ground truth labels regarding instances with different phone numbers forces us to tune parameters by only using advertisements that should be clustered (advertisements from repeated phone numbers), making it hard to minimize the false positives rate. This specific task could be enriched and better guided by incorporating domain knowledge from the law enforcement or human trafficking investigators in order to have true negative samples instead of the ones described in section 3.2.2.

5 Future Work

While the proposed entity resolution pipeline performs well in identifying sources within online escort advertisements written in Spanish, it could be improved by testing different variations such as a different ratio between positive and negative pairs in the contrastive learning training set and using a different approach for choosing negative samples that could also be guided and enriched by domain knowledge from experts.

Additionally, the pipeline should be evaluated on escort advertisements written in other languages to asses its performance even when the methodology does not depend on language based methods.

Lastly, one of the disadvantages of using the entity resolution pipeline proposed in this paper instead of methods that use Information Extraction frameworks based on regular expressions is the loss of interpretability from using a BLSTM as part of the framework. Using post-hoc interpretability techniques to explain which parts or characteristics of the text better explain the general aspects of each cluster would help in making this approach more interpretable without requiring manual validations.

References

Alvarez, M. and Burbano Acuña, D. (2017). Identifying human trafficking patterns online.

Alvari, H., Shakarian, P., and Snyder, J. (2017). Semi-supervised learning for detecting human trafficking. *Security Informatics*, 6.

Caulkins, J. P., Kammer-Kerwick, M., Konrad, R., Lee Maass, K., Martin, L., and Sharkey, T. (2019). A call to the engineering community to address human trafficking. 49(3).

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations.

Dubrawski, A., Miller, K., Barnes, M., Boecking, B., and Kennedy, E. (2015). Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking*, 1:65–85.

Giorgi, J., Nitski, O., Wang, B., and Bader, G. (2021). Declutr: Deep contrastive learning for unsupervised textual representations.

Ibanez, M. and Suthers, D. D. (2014). Detection of domestic human trafficking indicators and movement trends using content available on open internet sources. In *2014 47th Hawaii International Conference on System Sciences*, pages 1556–1565.

Kejriwal, M., Ding, J., Shao, R., Kumar, A., and Szekely, P. (2017). Flagit: A system for minimally supervised human trafficking indicator mining.

Kejriwal, M. and Szekely, P. (2017). Knowledge graphs for social good: An entity-centric search engine for the human trafficking domain. *IEEE Transactions on Big Data*, PP:1–1.

Konrad, R., Trapp, A., Palmbach, T., and Blom, J. (2016). Overcoming human trafficking via operations research and analytics: Opportunities for methods, models, and applications. *European Journal of Operational Research*, 259.

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

Nagpal, C., Miller, K., Boecking, B., and Dubrawski, A. (2015). An entity resolution approach to isolate instances of human trafficking online.

Neculoiu, P., Versteegh, M., and Rotaru, M. (2016). Learning text similarity with Siamese recurrent networks. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 148–157, Berlin, Germany. Association for Computational Linguistics.

Rabbany, R., Bayani, D., and Dubrawski, A. (2018). Active search of connections for case building and combating human trafficking. pages 2120–2129.

Ram, A., Sunita, J., Jalal, A., and Manoj, K. (2010). A density based algorithm for discovering density varied clusters in large spatial databases. *International Journal of Computer Applications*, 3.

The United States Department of Justice (2018). Justice department leads effort to seize backpage.com, the internet's leading forum for prostitution ads, and obtains 93-count federal indictment.

UNODC (2020). Global report on trafficking in persons.