

Persona Generation for Human Trafficking Victims

Nada Alghofaili, Joanna Cao, and Sampath Jayarathna

Department of Computer Science,
California State Polytechnic University
Pomona, CA 91768, USA

(nsalghofaili, jhcao, ukjayarathana)@cpp.edu

1. INTRODUCTION

Human trafficking is the third-largest criminal activity in the world, with millions of victims involved globally [1]. Every government worldwide has been fighting against this issue and not without a struggle due to the covert and discreet nature of human trafficking.

With the use of rising social media technology, people of all ages can create their own online accounts— thus generating a pool of potential human trafficking victims. Young children and teenagers especially are unaware of the presence of online predators and release their personal information to the public; making it easier for online predators to find vulnerable prey geared to their preference. Using online direct messaging platforms such as Kik, WhatsApp and Snapchat, these criminals can gain intimate access to these potential victims and gain their trust while hiding behind a screen.

This paper proposes a methodology to create personas; personas are typically created as an accurate yet fictitious individual enriched with demographical and behavioral details. The general purpose of personas is to interpret and synthesize descriptions of user segments, and in light of our research goal, which is to predict and uncover human trafficking incidents, personas can be refined through the analysis of data obtained from social media platforms, and be used to identify potential trafficking victims online.

The advancement of online communication technologies has resulted in these platforms becoming a major resource for human traffickers. According to THORNE, 63% of the victims recruited for prostitution were advertised online by their attackers [2]. Our objective is to generate automatic personas (synthetic profiles) to help identify potential victims of human trafficking activities as well as detecting traffickers online. We present the use of data

collected through social networks and other sources to generate personas. Ultimately we hope that the methodology can be extended to automatic persona generation and to use them with trafficker personas to help profile human traffickers.

2. BACKGROUND AND RELATED WORK

By analyzing how the online traffickers use the social media, these patterns can be identified and classifiers can be created to help counter these crimes. Through the assimilation of data from online classifieds and phone number analysis, exposing the activity of potential sex traffickers or victims as well as signs of sex trafficking has become possible [3]. Interestingly, the research on this topic is limited, and as a result, the feasibility of pinpointing online human trafficking occurrences warrants further investigation.

Obtaining first-hand user data for the personas that is statistically accurate can be difficult, which is where the application of social analytics [4] to obtain public user information in real time can be helpful. By analyzing those user's behavior we can provide an accurate framework for online personas that can be updated based on potential interactions of the users. In other words, by using these APIs to collect our data instead of using focus groups, surveys, or ethnography methodologies, we can automatically generate personas, saving both time and money [5].

With automatic persona generation as a viable option, government agencies combating human trafficking can profile multiple personas that represent both human traffickers and their victims, and use those profiles to investigate perpetrators and potential victims in the online world, stopping the crimes before they occur.

3. APPROACH

The methodology of creating a complete persona goes through two main phases. First, a basic persona is generated using the demographical information such as age, race, trafficking category and gender. A name will then be generated based on the indicated race. This will work as the “skeleton” of the persona, providing us with a framework with which we can add more personalized details pertaining to the persona obtained from the second phase. This second phase is the persona enrichment part where the data is collected from social media websites. The persona would be enriched with information such as a Twitter bio,

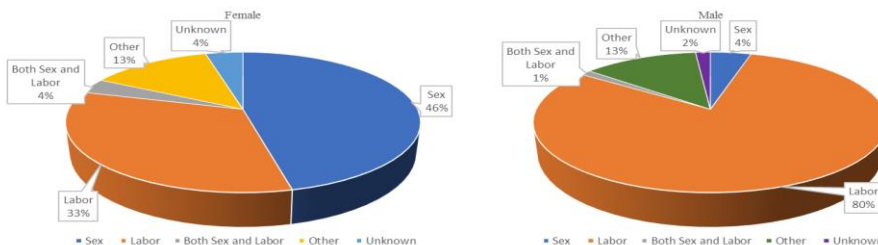


Figure 1: Victims Served by The Coalition to Abolish Slavery & Trafficking (CAST) Los Angeles, (Left) Female, and (Right) Male

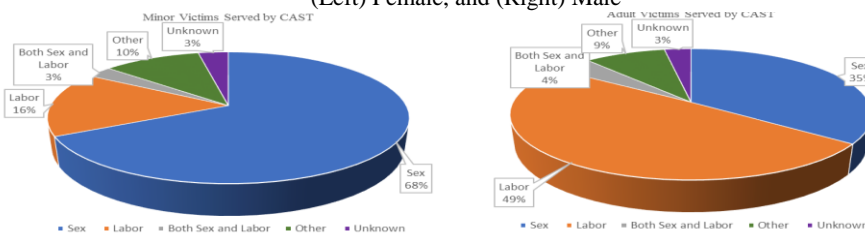


Figure 2: Victims Served by The Coalition to Abolish Slavery & Trafficking (CAST) Los Angeles, (Left) Minor, and (Right) Adults

sample Tweets, and other details such as mentions, retweets, etc.

3.1 Basic Persona Generation

To create the basic persona, we will be using data obtained from CAST. CAST is the Coalition to Abolish Slavery and Trafficking, and they are a Los Angeles based anti-human trafficking organization that helps rehabilitate survivors of human trafficking, raises awareness, and asserts influence over legislation and public policy surrounding human trafficking.

The dataset includes anonymous information of nearly 9,000 of their clients with open, waitlisted or previous cases of human trafficking, and provides metadata such as age, gender, race, trafficking category, etc. The demographics of a persona can be generated using this CAST dataset, and a name is randomly assigned from a simple dictionary of popular names based on the persona's race.

We summarize trafficking victim population served by the CAST-LA through snapshot reports (see Figure 1 and 2) which combines the data to provide a holistic picture of the activities of trafficking. It highlights key statistics on the trafficking population served, types of trafficking, and age. 58% of the total victims served were adults. Among the adult victims 49% were victims of labor trafficking; comparatively, 68% of minors were victims of sex trafficking. 77% of the total victims served were women. 80% of male trafficking victims were victims of labor trafficking, compared with only 33% of female victims. 13 victims were categorized as transgender. Of the total victims served, 65% were foreign nationals and 35% were US Citizens or Permanent Residents. This data was manually analyzed, and we were able to notice general trends and preferences in terms of who human traffickers tend to target and for what type of trafficking. For example, it was found that 32% of trafficking victims were of Hispanic, Latino, or of Spanish origin, with the second most trafficked race being African Americans at 18%.

Once this stage is implemented, a basic skeleton of a persona can be created by choosing the persona's race, age, trafficking category and gender, and then producing a name for the persona based on the race that was chosen (an example could be a 19 year old African American female victim of sex trafficking).

3.2 Persona Enrichment

A virtual LAMP (Linux OS, the Apache HTTP Server, the MySQL, PHP) server was set up to create the necessary environment for continuous, automated data collection. Since we are using Twitter to obtain our enrichment data, in order to capture the Tweets and parse them respectively, we are using the tools Phirehose and 140dev. All Tweets collected (along with their metadata, including any necessary information about mentions, URLs, and the original poster's profile information) are stored for later processing. During the process of applying the data from this database, the about section of the persona will be generated in a similar way to how the Twitter bio and sample Tweets are generated, and the persona's details are formed into a sentence format. The tweets collected, due to our choice in keywords input into 140dev, are similar to what traffickers look for the most when searching for victims online: those who are weak, vulnerable and in need of friends, money and support.

4. RESULTS AND CONCLUSION

4.1 Results

The figure 3 provides an example of the Persona Generation Form that we have created. The generation of the persona is initialized

The image shows a web form for generating a persona. At the top, there are input fields for 'Age', 'Gender', 'Trafficking Category', and 'Race', along with 'Generate Persona' and 'Reset' buttons. Below the form, the generated persona details are displayed under tabs for 'Persona Details', 'Twitter', and 'Other'. The 'Persona Details' tab is active, showing the name 'Sara', age '19', gender 'Female', trafficking category 'Sex Trafficking', and race 'Caucasian'. A circular profile picture of a young woman with red hair is shown. Below this, a 'Twitter Bio & Tweets' section displays a bio: '19 // heroin has the peace of god' and several sample tweets. The 'Other Interests' section lists 'Sports', 'Celebrity News', 'Fashion', and 'Travel Blogs'.

Figure 3: Sample Persona

by asking for indication of the personas age, race, trafficking category and gender. Once this is indicated, the persona will then be generated from the database of social media metadata, displaying the name of the persona, along with the information requested on the form. Following this general information would be samples of what the persona's Twitter profile would contain, such as a Twitter bio and sample Tweets. It would also provide other details concerning the persona, such as interests and dislikes.

In the future we hope to be able to automate the process of the skeleton persona generation possibly through the use of clustering tools, as well as automating the process of tweet collection, cleanup, and choosing the sample tweets for the persona. As of right now, the names of the generated personas are procured using only the race that is indicated. We plan on adding the indicated age as a factor in how our names are chosen as well.

5. REFERENCES

- [1] FBI, "Human Trafficking/Involuntary Servitude," May. 2016.
- [2] V. Bouché and Others, "A report on the use of technology to recruit, groom and sell domestic minor sex trafficking victims," 2015.
- [3] S. G. Jung, J. An, H. Kwak, M. Ahmad, and L. Nielsen, "Persona Generation from Aggregated Social Media Data," Proceedings of the, 2017.
- [4] UNODC, Global Report on Trafficking in Persons 2016. 2017.
- [5] M. Latonero, "Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds," Sep. 2011.