Using Similarity Measures to Detect Organizations in Online Escort Advertisements

Carl Edwards, Anthony Wertz, and Artur Dubrawski

Motivation

- Human trafficking is still a pervasive and global problem today
- 40 million people globally are involved in modern slavery [1]
- This activity has moved online in the form of social networking and online classifieds [2]
- The dataset consists of roughly 40 million advertisements from the escorts section of backpage.com from 2012 to 2017
- The usage of similarity measures can incorporate multiple modalities to allow for detection and monitoring of organizations

Similarity Measures

- **Text Similarity:**
  - Train fastText on corpus of advertisements
  - Generate paragraph embeddings of advertisements as average of constituent word vectors
  - Use cosine similarity between paragraph embeddings
- **Common-Feature Similarity:** Do two ads share some feature in common
  - Phone Number Similarity
  - Image Hash Similarity
  - Name Similarity
- **Face Similarity**
  - Face recognition pipeline implemented using DLIB:
    - CNN face detection
    - 5-point landmarking and face localization
    - Creation of face 128D face embedding using ResNet with 29 convolutional layers

Results

- Cosine similarity between unsupervised paragraph embeddings can be used to extract related ads in noisy text data.
- Additional similarity measures can be used to link organizations together amidst noise
- Multiple modalities can be used to improve the connection between ads
- Organizations can be monitored within single cities and on a country-wide level

References


Acknowledgements

I would like to thank Anthony Wertz and Dr. Artur Dubrawski, and the rest of the Auton Lab for their advice. I would also like to thank Rachel Burcin and Dr. Dolan for all their support. Finally, I would like to thank the RISS cohort for a wonderful summer and shared memories.

Future Work

- Incorporation of new similarity measures
- Further leverage visual information using background and foreground segmentation and matching