Modus Operandi, Discovering Human Smuggling Routes

Jan-Willem Tip, Cas Hortensius, Vera de Brouwer

Introduction

Identified issue:
UNODC: "report on human trafficking exposes modern form of slavery"

Problem definition :
- How are criminals operating?
- How can we make intelligence driven operations more efficient?

The goal:
- Develop method to extract Named Entities
- Develop tools for analyses support
- Improve police analyst process

Our research question:
Can we identify techniques to discover MO patterns in human trafficking court sentences?

Methodology

Steps
- Data collection

- Exploratory Data Analysis

- Filter entities from text with Named Entity Recognition

- Calculate Similarity between extracted routes

- Clustering based on similarity scores

Filtering of entities
Geolocations, Organizations, Non-Geolocations, Nationalities and our own added entity Vehicle

Similarity & Clustering
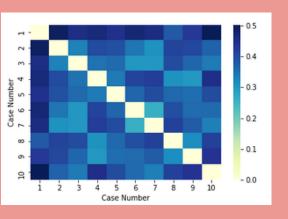Jaro similarity used for clustering Algorithm
Clustering Algorithm = Agglomerative Clustering
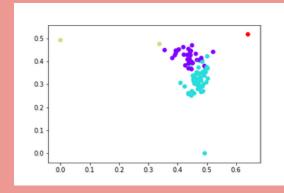
Results

Extracting routes from text

> een Vietnamese een Albanese door Groot-Brittannië een boot haven Seaport
> n IJmuiden naar Groot-Brittannië in Groot-Brittannië in Turkije C.O. Mar
> enstein'

Similarity scores

A heat map has been produced of the complete similarity matrix in order to gain insight into whether clusters could possibly arise. As can be seen from the heatmap (small part of the complete set), no clear clusters or associations can be demonstrated.

Clustering

The similraity scores were included in a cluster algorithm to determine whether clear clusters could be determined. As can be seen, apart from a few outliers, all documents can actually be grouped together in the same cluster. Therefore, no clear clusters can be separated from each other.

References

https://www.unodc.org/unodc/en/human-trafficking/human-trafficking.html

B.E. Turvey (2013), Modus Operandi, Encyclopedia of Forensic Sciences

Brouwer, P & Scholten, M, Joint Doctrine Publicatie 2 Inlichtingen, Defensiestaf

Conclusion

- Clustering on similarity shows no significance
- If a route is not present in the conviction, our 'route' becomes meaningless
- The potential is there, a lot more tweaking is required
  Watch out for implementing bias

Discussion

- Train model on context
  - Smuggling with boats → add seaport names, boat names, etc
- Look more into relevant words like 'te, en/of, etc' to find a significance
- Find way to find nationality of smugglers and victims