Vrije Universiteit Amsterdam

Master Thesis

A Clustering Method to Detect Mobile Organised Crime Groups

Author:
Floor Otsen

Supervisors:
Prof. Dr. Rob van der Mei
Sjanne Nap MSc.

Second reader:
Dr. Eliseo Ferrante

A thesis submitted in fulfilment of the requirements for the degree of MSc. Business Analytics

at the

Vrije Universiteit Amsterdam

August 20, 2020

Summary

Mobile Organised Crime Groups (MOCGs) are defined by the Dutch Police as a manifestation of internationally organised property crime; the offenders are only temporarily in the country where the crime is committed (for the purpose of this study - the Netherlands) and have no connection with that country. MOCGs are a complex and broad phenomenon, which currently takes a lot of manual effort for the Dutch Police to tackle, in which the results of the operation depends on the knowledge of the experts. To create additional insight into this phenomenon a clustering model is created on internal police data.

The cluster techniques that are applied to cluster the incidents of MOCGs are k-modes and greedy agglomerative clustering. The former method is an extension of the popular k-means and is applicable to be used with categorical data. It defines clusters based on the number of mismatching features between data points. For the latter method we have used the category utility (CU) function as goodness-of-clustering method. The CU function aims to maximise *both* the probability that two observations in one cluster have attributes in common *and* the probability that observations from different clusters have different attributes. For this method, the data first needs to be encoded, in which the data is represented in a numeric value for each feature or a binary vector. The features that are being used as input in the clustering models are the *crime category*, the *location type* and the *stolen goods*.

To evaluate the clustering methods, the adjusted rand index (ARI) and the adjusted mutual information (AMI) are computed for the predicted clusters and the clusters that are considered as 'true' in the validation set. The overall optimal number of clusters is determined by the performance evaluation metrics, and both metrics are considered equally important. The optimal number of clusters lies between 28 and 38 according to the performance evaluation metrics. The greedy agglomerative algorithm with numerical encoding performed overall the best on external validation metrics, and this was for 28 clusters. For the best model, the *similarity* of the incidents of the predicted clusters compared to the validation clusters (ARI) is on average 50% correct, and the *agreement* of the content of the incidents of the clusters (AMI) is a little higher, 57%.

The clusters of the validation set are visible in the results of the model, but often clusters of the validation set are merged in the results. Furthermore, incidents of the same suspect appear too often in different clusters. The results show that smaller clusters perform better than large clusters.

Preface

This thesis is written as graduation project for the Master Business Analytics at the Vrije Universiteit Amsterdam. The research was part of a six months internship at the Dutch Police, district Amsterdam.

I am very happy that I was able to perform my thesis with the Dutch Police and I would like to thank René Melchers for opening up a position in his team Business Intelligence & Quality (BI&Q). Furthermore, I would like to thank Sjanne Nap for introducing me within BI&Q, and also for being my supervisor within the police, for all her time, for helping me find my way and for giving me extensive feedback.

I would like to thank René Middag, one of the experts in MOGCs within the Dutch Police, for introducing me to the MOCGs theme, answering all my MOCGs related questions, discussing the results and for the exiting day work shadowing I had with him.

I am also grateful to Ruben van der Linden, for answering the data related questions and helping me with collecting the data.

I thank Rob van der Mei, for being my supervisor at the VU. I really appreciate the informal contact during the meetings, giving me guidance and feedback to develop and improve my thesis. Finally, I am grateful to Eliseo Ferrante for his involvement in the project as the second reader.

Contents