

Predicting Troop Betrayal using Anomaly Detection and Sentiment Classification

Arnav Singh, Atharva Prashant Patil, Lakshya Bhatnagar,
Rushil Vijaykumar Patel, Satyam Sharma

September 2024

1 Introduction

We have designed a model which will predict the likelihood of betrayal of a particular soldier of the troop. Some of the important factors were selected to quantify the factors such as financial incentives, personal grievances, social participation, ideological differences.

2 Hypotheses

1. It is very unlikely that a group of soldiers together would betray because if that is the case then the anomalies detected by the algorithm might be detecting the people who would not betray Xernia.
2. It is also assumed that we can understand the language spoken by the soldiers of our troop so that we can run the sentiment classifier algorithm.
3. Xernia is governed by a totalitarian government; thus, it can access private data of any soldier.

3 Factors Selected

The following are the factors selected along with an explanation of why they were selected and how to quantify them as features.

3.1 Financial

Soldiers with lower financial standing may be more likely to succumb to financial incentives offered by the enemy. The financial standing of a soldier can be measured through following factors:

1. Salary
2. Net worth
3. Debt

3.2 Respect

In any civilization the respect of an individual is greatly determined by their professional life:

1. Rank
2. Honours: The honours awarded in the army can be ranked according to their prestige. The sum of those ranks can be treated as a feature.

3.3 Personal Factors

1. Age: Soldiers of certain age may be more likely to betray.
2. Personal Calls: percentage of number of calls received by a person from known contacts

4 Workflow

The workflow will consist of three different sub parts.

4.1 Anomaly Detector

This algorithm would be provided with all the features as the input. It may classify the soldier as an anomaly. The output would be either 0 or 1. This output also will be fed into the input in the neural network.

4.2 Sentiment Classifier

The suspicious conversations of every soldier will be recorded through their phones and fed into a Sentiment Classifier. The sentences would be padded before being fed. The algorithm classifies the input sentences provided into two classes: will betray and will not betray. The algorithm would incorporate the use of word embeddings and LSTMs. GloVe vector embeddings are used.

4.2.1 Data Collection

The phone of any soldier is constantly hearing the conversations. When it detects one of the predefined trigger words (the set of the words will be created using most frequent words occurring in the calls of the previously identified betrayers), the phone starts to record the audio till the phone stops detecting the trigger words for span of 5 minutes.

4.3 Computing Strongly Connected Components of a graph

The social media activities of the soldiers would be tracked. SCC of previous betrayers would be identified. If the SCC of a soldier matches with this SCC, it will be flagged and fed as an input feature to the Neural Network that will predict their loyalty.

4.4 Neural Network

The neural network will take all the features as the input. The features will be provided to the neural network after normalization to speed up learning. Methods such as Adam Optimization, regularization and drop out may be implemented after viewing the errors obtained in training and cross validation sets.

The output of the Neural Network will be between 0 to 1 where higher score mean higher chances of betrayal.