
Identification of Safest Path using Spatio-Temporal Analysis

1 Problem Statement

In this project, we propose a method to find the safest path between two locations, based on geographical models of crime intensities. We use the spatio and temporal modeling techniques to identify the location and relevance(based on the date) of the crime. We consider the police records and news articles as the basis for our calculations. It is essential to consider news articles as there is a significant delay in updating police crime records. We address this problem by updating the crime intensities based on current news feeds. Based on the updated crime intensities, we identify the safest path. Our model would even inform the user of crime sprees and spurt of crimes in a particular area thereby ensuring that user avoids these crime hotspots.

2 Motivation

In today's society criminal activities are on the rise. Newspapers each day are full of news articles shrieking about crime incidences from all corners of the world. Governments all over the world are spending millions trying to curb this menace. We intend to come up with a way by which one can ensure that he travels from one place to the other by the safest route possible. This data mining paradigm we have developed mines essential information from police FIR records and Newspaper articles to provide the user with the most up to date information of the safest path and crime hotspots of any region.

3 Datasets

The police FIR records shall be obtained from the respective police websites. The FIR's registered at a particular police station would be considered as the prior crime intensity of that police station. We intend to extract the crime records starting from 2011 and exponentially decrease the relevance of the crime article as time for current day increases.

News articles are obtained from the various News paper's websites using web crawlers. These are essential to keep our model up-to-date with the current affairs. The news articles shall be categorized as crime/non-crime and their location and time shall be extracted through spatial and temporal analysis to identify the location and time of the article.

4 Procedure

4.1 Categorization of article as Crime/Non-Crime

We will use Latent Semantic Analysis to classify the article as crime/non-crime [1]. We convert the articles to a term-document matrix i.e. for every document we record the frequencies of the

words occurring in these documents and then apply Singular Value Decomposition to obtain a vector representing the article[4]. Thereafter, the article can be classified as crime/non-crime using k-means algorithm.

4.2 Identifying Location and Date/Time of Crime

Spatio-temporal analysis would be implemented to obtain the heat-map of the location and date/time of the article[3]. We will use YAGO2, an extension of the YAGO knowledge base, in which entities, facts, and events are anchored in both time and space. YAGO2 is built automatically from Wikipedia, GeoNames, and WordNet. [2]

4.3 Mapping Crime Intensities

For this we identify the police station in the region and assign it weight based on the following parameters:

1. Crime Frequency of the region
2. Time/Date of each crime in the region

4.4 Identification of Safest Path

Once the crime intensities are known ,we can convert the safest path problem to the shortest path problem by modeling the edge weights as the sum of crime frequencies of the two connecting nodes. Now that we have a graph with well defined positive edge weights,we can apply Dijkstra’s algorithm to identify the shortest path which is the safest path here.

5 References

References

- [1] S. T. Dumais. “Latent Semantic Anlaysis”. In: *Annual Review of Information Science and Technology* vol. 38 (2004), pp. 188–230.
- [2] Johannes Hoffart et al. “YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia”. In: *Artificial Intelligence* vol. 194 (2010), pp. 28–61.
- [3] J. Ratcliffe. “Crime Mapping: Spatial and Temporal Challenges”. In: *Handbook of Quantitative Criminology* (2010), pp. 5–24.
- [4] G.W. Furnas T.K. Landauer R. Harshman S. Deerwester S.T. Dumais. “Indexing by Latent Semantic Analysis”. In: *American Society of Information Science* (1990), pp. 391–407.