Initial Plan

Title: Identifying Crime Hotspots using Twitter

Name: Redmond Todd-Bennett
Student Number: c1124856
Supervisor: Dr Peter Burnap
Moderator: Dr Xianfang Sun
Module: CM3203- One-Semester Individual Project
Credits: 40

Project Description

Data is constantly being created and this has led to the emergence of big data sets which are so large they are becoming difficult to process so a current challenge is what value can be derived from them. Big data analysis can be very useful to see if we can extract any potentially useful trends or patterns. Examples of big data sets include social media platforms Facebook and Twitter.

In my project I will specifically analyse references to crime and insecurity from Twitter. Previous work carried out in this area includes scraping Twitter looking particularly at crime-related semantics and natural language of tweets to develop a model which can predict crime happening. Away from social media and big data, spatio-temporal analysis of crime scenes has been visualised to detect certain hotspots of crime using geographical information systems (GIS).

The next step I will take in this project is to be able to deduce that a criminal event has happened from Twitter data. By analysing the natural language of tweets for specific crime-related terms or phrases I will be able to evaluate a crime has occurred. Where possible, leveraging the location metadata which is linked with a tweet, I will plot a geospatial distribution of these crime events to show where there are particular hotspots of crime. Finally, I will perform a statistical analysis to show whether there is a correlation between actual reported crime and those crimes that have been identified from Twitter data. Using Pearson chi-squared test I will deduce whether there is some correlation or if there are any differences whether they occurred by chance.

In order to support solving this problem I will need to produce a web-based system which uses Google Maps API to plot and visualise locations of criminal events from Twitter data.

Project Aims & Objectives

Core Deliverables
- Source Twitter data to use for analysis and modelling in proposed system

- Identify key terms or hashtags in Twitter data referring to crime and link together to deduce particular criminal event has happened.
    - Focussing on a particular spatial area, I will take a subset of the data and identify specific crime-related terms and phrases in the natural language or hashtag of tweets. Then by applying these rules I can classify other tweets in the full data set.
- Develop system to visualise the distribution of criminal events from Twitter data onto a map using Google Maps API (utilising heat-map or markers functionality)
- Visualisation of particular hotspots of criminal activity based on results of the above deliverables
- From results undertake a statistical analysis to see if there is a correlation between actual reported crimes and the crimes identified from Twitter data

Optional Additional Deliverables
- Develop system further to split crime hotspots distribution by crime type

Work Plan

My approach will combine the documenting of the project report with designing and implementing the proposed system solution to the problem.

Week 1
- Write Initial Project Plan
- Research and experiment with Google Maps API which I will use to visualise crime data

Week 2
- Document Project Introduction and Background by researching further into similar work which has already been carried out in this area
- Identify potential other tools/resources which will help to meet project aims and objectives
- Source and analyse Twitter data set (provided by Cardiff School of Computer Science & Informatics)

Week 3 - Week 4
- Document Project Approach
- Begin designing proposed system referencing how the system will be developed and the interlinking of tools to meet the project aims
- Week 4 - Supervisor Review Meeting to discuss project progress and any alterations which may be required at this stage

Week 5 - Week 9
- Document system design section of report

- Develop system implementation by creating solution which will visualise locations of identified tweets which have a reference to crime
- Ongoing System testing to ensure it is functioning in line with project aims and objectives
- Week 8 - Supervisor Review Meeting to discuss project progress and any alterations which may be required at this stage

Easter Recess Week 1 - Easter Recess Week 2
- Document Implementation picking out critical or innovative code which has helped to meet project aims
- Analyse results and findings identifying particular hotspots of crime based on geospatial distribution
- Carry out statistical analysis (Pearson Chi-Squared Test) to see correlation between actual reported crime and the hotspots identified in the developed system

Easter Recess Week 3
- Document Results section and statistical analysis considering most appropriate data representation
- Research and document potential future work in this area and which could take this project further

Week 10
- Document conclusions based on the output of the results and how the project has met the aims established in this Initial Project Report
- Reflect on and document learnings from the Project identifying relevant strengths, weaknesses and improvements which could be made

Week 11
- Proof reading and completion of outstanding sections of Project report
- Contingency time to complete any revisions to the Project report

Week 12
- Submit Project report and relevant associated materials