Fundamental Concepts in Data Insight:
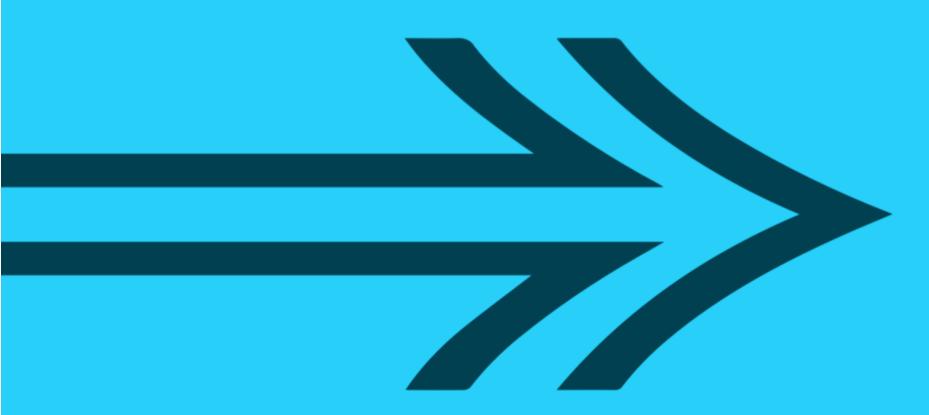
Policing with Data & Case Studies in UK Policing

Fundamentals for a General Audience

Policing with Data & Case Studies in UK Policing

- Case Studies in Problems
    - Challenges in Data Collection
    - Case Study: House Burglery & Data Siloing
- WB. Recording a Crime
    - Review: Recording the Crime
    - Review: Intelligence Gathering
    - Review: Arrest
- Challenges in Data Quality
    - Problem: Structure
    - Problem: Field Structure
- Problems
    - Problem: Uniqueness
- Challenges in Modelling
    - Problem: Behavioural Change
    - WB. Future = Past?
    - WB. Criminal Network Analysis
        - Review: Problem: Network Analysis
    - Problem: Priotisation of Offenders
        - WB. Factor Analysis
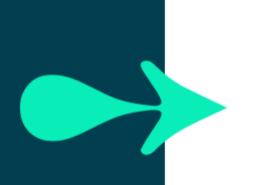
- Challenges in Research & Experimentation
    - Problem: Getting Data
        - Problem: Unifying and Annoymising Data
        - Projects: Europol EPRIS
    - Problem: Decision Quality
    - Break Exercise
- Case Studies in Solutions
    - Case Studies: Supervised Learning
        - for Interview Crime Lists
    - WB. Supervised Learning
    - Case Studies: Unsupervised Learning
        - for Early Detection of Crime Series
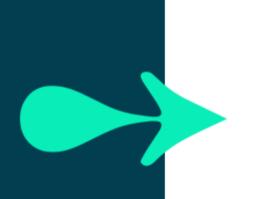   - WB. Unsupervised Learning

Case Studies in Problems

Challenges in Data Collection

Case Study: House Burglery & Data Siloing

WB. Recording a Crime

WB. Recording a Crime

WB. Recording a Crime

Review: Recording the Crime

- House Burgled

    - Crime Reported
    - Command & Control DB

- Officer Attends Crime Scene

    - Takes Details
    - Crime Recording System
        - Command & Control not sync'd

- CSI attends Crime Scene

    - CSI updates Forensic Reporting System (not sync'd)

Review: Intelligence Gathering

- National Intelligence Model
    - prescribes a tasking-and-coordinating session
    - http://www.intelligenceanalysis.net/Legal%20-%20Gwent%20NIM%20policy.pdf

- eg., prescribe Stop & Search in areas
    - Stop & Search System
    - to Gather Intelligence, eg., "Saw X near Y doing Z"
        - Intelligence Database
        - National Grading model
            - rating of intelligence (/5)
            - who it can be shared with, quality, etc.

Review: Arrest

- Offender Arrested
    - Custody System

- Crime Colved

    - Crime Nominals System
        - suspects, victims, convicts
    - Crime MOs registered

- National Databses Sync'd

    - https://en.wikipedia.org/wiki/Police_National_Computer
    - https://en.wikipedia.org/wiki/Impact_Nominal_Index

Challenges in Data Quality

Problem: Structure

- structured data
    - fields, etc. in the systems
- semi-structured data
    - field report forms -> automated ingest
- unstructued data
    - raw police officer notes

Problem: Field Structure

```python
pd.read_csv('crimes_freetext.csv', index_col='CRN')
```

CRN	Locations	MO Description	MO notes	Suspect Description
ABC-1	leigh on sea, essex	WINDOW 3f SMASH NA NA NA NA	suspect seen fleeing scene	two susepcts c. 6ft brown hair
ABC-2	leagh on see	DOOR 5f FORCED NA NA ALL NA	only jewlery taken	6f and 5ft 6 seen together fleeing
ABC-2	leigh-on-sea	REAR WINDOW NA NA NA	took all cash	brown hair

Problems

- locations
  - free text fields
  - hard to parse names (comma, dash..)
- MO description
  - populated by forms
    - repeated "NOT KNOWN"
- MO notes
  - free text
  - NLP to parse MO notes "Mosaic Project" (UK)
    - NLP company behind project at a loss on police data!
- suspect descriptions
  - free text
  - 1m64, 5f5, etc.
  - do the descriptions relate to a single offender, or multiple?
    - "how many nominals?"

Problem: Uniqueness

- duplication

  - unique reference number (URN) per person?
    - no, operates create additional records "for convenience"
  - https://www.experian.co.uk/business/glossary/golden-nominal/

- "golden nominal" -- one identifier

  - free text fields need matching
    - phonetic matches?
    - complex text matching: lucene
    - https://lucene.apache.org/core/6_1_0/analyzers-phonetic/index.html
    - https://en.wikipedia.org/wiki/Soundex
  - quality?
    - 98% per individual record match
    - 85% in the aggregate
      - compared to human performance

Challenges in Modelling

Problem: Behavioural Change

- what are the statistical presumptions of predictive analysis?
    - historical distribution matches future distribution
- offenders learn from other offenders
    - prison
    - MO changes

WB. Future = Past?

WB. Criminal Network Analysis

WB. Criminal Network Analysis

Review: Problem: Network Analysis

- criminal network analysis

  - https://www.ibm.com/security/intelligence-analysis/i2/law-enforcement
  - overt = public social network, friends
  - covert = contact events hidden from view
    - connections only visible when suspects interact with police systems
      - criminals work together on crimes
      - two nominals against one crime
        - aside: link to stop & search, other dbs
    - social network algs don't work on covert networks
  - who do you target in a network?
    - degrees of separation
      - exponential number of nodes included
      - low-level crimes = 2deg
      - high-level = maybe 6deg, ie. gloabl

- consider organized crime network

  - how do you disrupt?
  - target by node centrality?
    - eg., betweenness
    - but centrality is often cooincidental
      - ..in covert networks!

Problem: Priotisation of Offenders

- very few single offenders
  - vast majority repeats
- automated process considers:
  - MO, spatial & temporal factors
  - high priority in overlap
- how do you target?
  - aside: targeting can be simple as sending letter, "we are watching"

WB. Factor Analysis

Problem: Priotisation of Offenders

- typical factor analysis, (run monthly?), risk score:
    - https://www.gov.uk/guidance/risk-assessment-of-offenders
    - l2 offender
        - cross border, per crime per border
    - offten priority
        - high crime area, live
    - live priority
        - hime crime area, visit
    - community saftey
        - weapons, ...
    - etc.
    - (external: beat area crime)
- resource & task modelling
    - how do you get police in right place at right time?
    - "person flows" through built-up areas
        - criminal flows, enviromental modelling

Challenges in Research & Experimentation

- goal: find operationally effective methods

Problem: Getting Data

- excess of 200 databases
    - joins taken in excel
    - some systems cannot be automated
    - built for *reporting* not insight
    - legal data retention rates
- national dbs often require special perm. to access
    - national systems often uniform (home office, national police dbs, are uniform)
    - one-off logins: login, export, logoff
- locals all different
    - some regional colab, "7 force colab"
    - https://www.theguardian.com/government-computing-network/2012/feb/03/police-northgate-athena-it-framework
    - https://www.bedfordshire.police.uk/information-and-services/About-us/Seven-Force-Strategic-Collaboration
- multi-agency data
    - data sharing agreement
        - some agencies are reluctant to share

Problem: Unifying and Annoymising Data

- unifying data across dbs,
  - universal id for *a crime*
    - very hard!
  - associating: case data, witness statements
    - against crimes, nominals, ANPRI, ...
  - databases (+1m/db/yr)
    - stop and search, crime, finance
    - custody, anpr
    - nominals (suspects, victims,...)
    - command & control, property
- can you annoymise data?
  - pseudo-anon requires registering/legal

Projects: Europol EPRIS

Europol's website does not have any current references to the Piracy database but has added the Travellers Database, which contains information on 34.000 "foreign fighters" (EUObserver, 2016). To increase the efficient exchange of police records for criminal activities that fall outside the scope of Europol, several European Member States have successfully lobbied for the creation of EPRIS (Focant et al, 2012).

Administered by Europol, the mandate of EPRIS is to provide police with a quick overview if and where police records on an individual can be found in Europe. As such Europol is developing EPRIS as an INDEX, which allows police forces to search in each other's databases. The result will be 'hit'/'no hit'; in the case of a 'hit' the querying members state police department can then go through the proper judicial channels to gain access to the information (Jones, 2011).

Problem: Decision Quality

- capturing investigative thought process
  - eg., crimes occuring over 3mo
    - breaking into vans
    - taking sat navs
    - 30+ crimes for one group
    - single investigator
      - how do they share their knowledge?
- analyst end-user interfaces and visuals
  - current systems often very manual, but high-quality
    - consider eg., court room use
  - "analyst workbench"
    - https://www.ibm.com/uk-en/products/i2-analysts-notebook -- £8k!!!!/user
  - visualization extremely important tool
    - most UK analysts use Excel
      - not sufficient for policing
- the system itself will never be used in court
  - so lower evidence-base required
  - TRL framework: TRL 5/6
  - https://en.wikipedia.org/wiki/Technology_readiness_level
- serious and organized crime "ecosystem"
  - http://old.heuni.fi/material/attachments/heuni/papers/6Ktmwqur9/HEUNI_papers_26.p

Break Exercise

- What are problems?
- How would you solve them?

Case Studies in Solutions

Case Studies: Supervised Learning

for Interview Crime Lists

- Suspects may find an advantage in admitting many crimes at once
    - concurrent/reduced sentence vs. sequential
- Problem: find related crimes
    - analyst time constrained
    - examine current crime, find some matches
        - interviewer (officer) will go thru with suspect
- Possible Solution: blackbox technique
    - train on historical cases
    - beat/crime/postcode/basic-command-unit (BCU)
        - https://en.wikipedia.org/wiki/Basic_command_unit
    - extant system
- Results of modelling
    - "accuracy" > 75%
    - 15min of computation time
    - cross-reference with crime-nominals-system

WB. Supervised Learning

WB. Supervised Learning

Case Studies: Unsupervised Learning

for Early Detection of Crime Series

- Are crimes linked?
    - Can we detected it?
- Possible solution: unsupervised technique
    - clustering
        - last 3mo of crimes
            - NLP/parse free text for keywords
        - "self-organizing map"
            - https://en.wikipedia.org/wiki/Self-organizing_map
        - cluster size = quality of cluster
            - 7 crimes good-quality
            - 40, bad
    - ie., basically clustering / factor analysis
        - determines crime similarity scores

- produce list of crimes from clustering
  - eg., MOs:
    - entry = rear, side
    - feature = window, door
    - feature type = fixed, roller
    - method = climbed, smash
    - rooms = all, one, up, donw
    - search type = tidy, untidy
    - unique id, name, dob,...
  - and thef., suggest *all* of the above for future crimes!
    - requires analytical judgement
- this is not *evidence*
  - it is grounds for evidence collection!
  - false positives, etc.

WB. Unsupervised Learning

WB. Unsupervised Learning

WB. Unsupervised Learning

Appendix

DB Case Studies

Project: UK DB Interoperability

The U.K and Germany are making significant investments to update and increase interoperability between different police databases. The U.K has a federated law enforcement structure, which historically has provided local police forces autonomy to buy and implement their own policing technologies and has led to the proliferation of different databases and database structures.

Project: UK DB Interoperability

The Home Office is spearheading by the National Law Enforcement Data Programme (NLEPD), under which it aims to create the National Law Enforcement Data Service (NLEDS) , connecting the databases of the different U.K police forces into one centralized system.

Project: UK DB Interoperability

The national ANPR database receives around 50 million ANPR 'reads' a day (Police.uk, 2019; Surveillance Camera Commissioner, 2016: 23). IBM has been awarded the £12,000,000 contract to assist the NLEPD with the transformation of the existing systems into the NLEDS (Home Office, 2016B).

Other Studies and Links

Case Study: Valcri Analytics Projct

- http://valcri.org/

Case Study: Accenture West Midlands Data Insight Project

- https://www.accenture.com/gb-en/case-studies/cloud/west-midlands-police
- https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution