Introduction: PIXLES

PIXLES is an interdisciplinary knowledge platform and research consortium with a focus on current societal dilemma’s and developments in the context of privacy, information exchange, law enforcement & surveillance. The core value of PIXLES is to safeguard privacy while enhancing security.
Overview session

1. New and emerging data sources for spatio-temporal analyses: Opportunities and threats for big data policing
   Thom Snaphaan (speaker); Wim Hardyns

2. Social capital variables at the neighbourhood level as predictors in a predictive policing model
   Anneleen Rummens (speaker); Wim Hardyns

3. Automated Suspicion: Offender-Based Predictive Policing in the Age of Big Data
   Bob Rigo (speaker)

Content

I. Introduction
II. Big data
III. Big data policing
IV. Scoping review: big data in environmental criminology
V. Conclusion and discussion
I. Introduction

The best predictor of future crime is ... ?

PRIOR CRIME

Big data policing is more than predicting the future (≠ predictive policing)

Is big data policing another type of policing? Next to, among others, intelligence-led policing (ILP), problem-oriented policing (POP), community oriented policing (COP), hot spot policing

More like a paradigm shift (which is not limited to the police)

Although big data predominantly have been used for crime prediction purposes, the use of big data is much broader than that

Big data as one type of research data

Crime prevention, policy

II. Big data

Dissecting the catch-all term

Innovative data sources

Innovative data collection methods

Innovative data processing and analysis methods
II. Big data

Definition and characteristics

Myriad of definitions of big data
→ Regularly described based on their characteristics

3Vs:

- Volume: enormous quantities of data
- Velocity: created and processed in or near real-time
- Variety: structured, semi-structured or unstructured

Kitchin & McArdle (2016):
"... the key boundary characteristics of Big Data, which together differentiate it from 'small data', are velocity (both frequency of generation, and frequency of handling, recording, and publishing) and exhaustivity" (p. 8) (captures an entire system, rather than being samples).

II. Big data

Two types of data

Two distinct types of data:

- Data of active solicitation: made by researchers for research purposes
→ 'made' data

- Data of passive solicitation: collected for purposes other than research, and researchers do not have any influence on the data collection and data processing
→ 'found' data

Source: (Chen, Ma, Suresh, Liu & Wang, 2016; Connelly, Playford, Gayle & Dibben, 2016)

II. Big data

Taxonomy of data sources in the social sciences

Data sources in the social sciences

Experimental

Observational

Automated

Volunteered

Retroactively user-generated content

Proactively user-generated content

Interaction data

Scan data

Sensed data

Data generated by digital devices

Automated surveillance

Administrative data

Data generated in digital devices

RTI data

Telematics data

Social media data

Inspired by: (Jadeja & Issa, 2018; Kitchin, 2013; Thakuriah, Tilahun & Zellner, 2017)
II. Big data

**Taxonomy of data sources in the social sciences**

**Directed data:**

Volunteered data (proactively and retroactively user-generated content):

- Yelp
- Facebook
- OpenStreetMap
- Twitter

Automated data (automated surveillance, data generated by digital devices, sensed data, scan data, interaction data):
III. Big data policing

**Black, blue, bright and no data**

Ferguson (2017): The rise of big data policing

- Black data: distortions of race, transparency and law
- Blue data: (internal) policing data
- Bright data: big data identifies the risk but not necessarily the remedy, and just because big data policing identifies the risk, does not mean every crime problem needs a policing solution
- No data: data holes

---

III. Big data policing

**A quick look at the relative importance**

Web of Science: today's premier academic database

Search: ("Big data" AND Policing) in Topic for the years 2005-2018

---
III. Big data policing

**Today's applications of new technologies**

Survey (n=46) regarding the use of new technologies by police forces and law enforcement agencies in 11 different countries (EU + Australia)

(Eustas & Vergouw, 2015)

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insufficient funding</td>
<td>Organizational</td>
<td>40%</td>
</tr>
<tr>
<td>Inadequate technology</td>
<td>Technological</td>
<td>30%</td>
</tr>
<tr>
<td>Lack of available data</td>
<td>Technical</td>
<td>30%</td>
</tr>
</tbody>
</table>

IV. Scoping review: big data in environmental criminology

**Evolution of the number of included studies**

![Graph showing the evolution of included studies](chart1)

IV. Scoping review: big data in environmental criminology

**Evolution of the number of data sources used**

![Graph showing the evolution of data sources](chart2)
IV. Scoping review: big data in environmental criminology

**Data sources used**

- BIBLIOMETRICS
- GOOGLE SCHOLAR
- SCOPUS
- WOS
- PUBMED
- SCIENTIFIC ABSTRACTS
- SCIENCEdirect
- JSTOR
- SCIENCE in students
- GOOGLE NEWS
- GOOGLE IMAGE

**Strengths**

- Real-time temporal and fine-grained spatial character ➔ time and place-specific dynamics
- Frequently open-source availability ➔ no or low cost
- Significant improvement of prediction models
- Data available for places where these were previously not available
- Unobtrusive measures potentially compensate several biases and errors present in data collected by researchers for research purposes ➔ have the potential to better describe social phenomena (note: equivocal views, cf. weaknesses)
- Big data enable new data processing methods (e.g., NLP)

**Weaknesses**

- Larger measurement error, because one may be unaware of how distilled variables from big datasets were measured ➔ need to infer the meaning of the data (note: equivocal views, cf. strengths)
- Sampling and selection bias: availability and 'digital divide'
- Does not always provide opportunities for longitudinal research

**Discussion**

- Technology is value-free (neutral)
- Technology is the means, it is not the end goal
- Methodology is fallible
Conclusion

(1) It takes two to tango
Cooperation and communication between computational (social) scientists and domain experts. For criminology: large opportunities for collaboration between or integration of ‘analytical criminology’ (scientific realism) and computational/social science
Also applies to policing!

(2) Framework to assess the quality of data sets and methods (cfr. TSE)

(3) FAT: Fairness, Accountability, Transparency in research and practice

Bibliography


Making the most of the available data: Social capital variables at the neighbourhood level as predictors in a predictive policing model

Anneleen Rummens
Wim Hardyns

17 September 2019
Predictive policing in the big data era
CPS international conference ‘Street policing in a smart society’

Contents
I. What is predictive policing?
II. (Big data) sources for predictive policing
III. Social capital research in Ghent: SWING and SCAN
IV. Predictive modelling using SWING
Conclusions

I. What is predictive policing?

- Predictive policing: “the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity” (Ratcliffe 2014, p. 4)
- Recent development: use of big data and predictive analysis in criminology
- Evolution to small units of analysis (street segments, grids) and complex machine learning methods
- More dynamic than methods such as hotspot analysis
I. What is predictive policing?

Main objectives of predictive policing:

• Short-term:
  - Use available resources more efficiently
  - Proactively target crime

• Long-term:
  - Cost-efficiency
  - Decrease crime rate

II. Predictive policing using the SWING data

  - Extensive study of relationship between social capital, crime and health
  - Collected data on among others social and physical disorder, social trust, informal control, …

• Doctoral research at Ghent University: Performance of different data sources and methods for predictive policing?
  - Retrospective analysis: predicting crime for a period that has already happened, so that we can compare the predicted zones to the real crime zones
  - Using the SWING data, a predictive model for home burglary in Ghent was created
II. Predictive policing using the SWING data

Including the SWING variables improved our base model (only crime history variables) with 1% for both recall and precision (a limited increase).

- Recall: 20.18% (% incidents correctly predicted)
- Precision: 17.00% (% predictions which are true events)

N incidents = 122
N predictions = 100

Advantages and disadvantages of using social capital variables in predictive policing

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Provides information on residents’ concerns</td>
<td>• Difficult to acquire dynamically</td>
</tr>
<tr>
<td>• Provides inside information</td>
<td>• Can be misleading for predictive purposes</td>
</tr>
<tr>
<td>• Social capital has been shown to have a strong relationship with crime</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion: Social capital variables (as collected by the SWING study) have low predictive value, but can be useful as supporting variables.

III. Selecting variables for predictive policing

**PREDICTIVE POLICING VARIABLES**

- Improves prediction of crime events
- Dynamic data

Examples:
- Crime history variables (previous incidents, time since last incident, ...)
- Ambient population (mobile phone data)

**SUPPORTING VARIABLES**

- Provides context
- Can help explain why certain places/times are at risk

Examples:
- Calls to emergency services
- Social data
III. Selecting variables for predictive policing

- One of the main characteristics of predictive policing, is its use of (big data) sources
  ⇒ More advanced systems even provide real-time integration

- New technologies open up new possibilities: ANPR, mobile phone data, ...

- Some applications are also looking at civilian inputs, e.g. ’citizen science’ apps
  ⇒ This kind of data can also be of interest in the context of predictive policing, for example, with regards to fear of crime, well-being, social control, cleanliness, ...

Such an approach could also help to make data more dynamic, thus increasing its predictive value.

Conclusions

- Making the most of predictive policing also requires making the most out of the available (big) data
  - Predictive vs supporting variables, e.g. social capital data
  - Use of big data & new technologies

- General guidelines for selecting variables for predictive policing, but specific context is important
  ⇒ Exploratory analysis needed before implementing predictive policing
  ⇒ Test different configurations

- Important questions to ask:
  - Which data resources are available?
    - Which data could be interesting to include?
      • for prediction?
      • for support?
Automated Suspicion: Person-Based Predictive Policing in the Age of Big Data

EU data protection law in a privacy-friendly mass surveillance society

Bob Rigo
+32 9 264 69 37
Bob.Rigo@ugent.be

“"To what extent are person-based predictive policing tactics legitimate in light of what EU data law is and ought to be in a privacy-friendly mass surveillance society?"
To what extent is person-based predictive policing legitimate in light of art. 11 LED?

What is person-based predictive policing?

- Predictive algorithms
- Algorithmic decisionmaking
- Prediction-led policing business process

Strengths

- Effective management information system
- Amplifications of prior surveillance practices
- New patterns
Some concerns ...

Data Analysis:
- False positives / Echo-chambers
- Opacity / Black box
- Police operations
- Automation bias
data protection as a fundamental right

- Article 8(1) (‘the Charter’)
- Article 16 TFEU (the Lisbon Treaty)
- 2008 Framework Decision, replaced by:
- Law Enforcement Directive

“Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her, to be prohibited unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller.” Art. 11(1) LED

“Member States shall provide for a decision based solely on automated processing, including profiling (…)”

- Predictions (vs. decision)
- Some form of automated processing (e.g. data mining)
“Member States shall provide for a decision based solely on automated processing, including profiling (…)” Art. 11 LED

- What constitutes a decision?
- Process matters
- Stages in the algorithmic process

“Member States shall provide for a decision based solely on automated processing, including profiling, (…) ” Art. 11 LED

Solely

- without any human intervention?
- Ambiguous

“Member States shall provide for a decision based solely on automated processing, including profiling, (…) ” Art. 11 LED

What constitutes human intervention?

Nominal (human rubber stamping) or meaningful human intervention?
"Member States shall provide for a decision based solely on automated processing, including profiling, (...)”  Art. 11 LED

Unauthorised law-making by Art. 29?

- “not any human involvement”
- “with an influence on the result”
- “meaningful”

Adverse legal effect (↔ GDPR)

- No definition
- Discriminatory effects?

Significant effect

- Less clear
- “substantial”
“(...) prohibited unless (...) appropriate safeguards for (...) at least the right to obtain human intervention on the part of the controller.” Art. 11 LED

• Human intervention?
• Automation bias

“(...) prohibited unless (...) appropriate safeguards for the rights and freedoms of the data subject (...)” Art. 11 LED

Some clarity?

• to express point of view
• to obtain an explanation
• to challenge the decision

“(...) prohibited unless (...) appropriate safeguards for the rights and freedoms of the data subject (...)” Art. 11 LED

However,
• Recital/Directive ↔ art. 15 GDPR?
• Trade secrets
• Need for flexibility (recital 44)
• Black box
“(…) prohibited unless (…) appropriate safeguards for the rights and freedoms of the data subject (…)”  
Art. 11 LED

Other safeguards?

Determinism vs. Probability
Two-way-mirror

To what extent is person-based predictive policing legitimate in light of art. 11 LED?
• Human intervention?
• Explainability?

• Legal nature
• Vague concepts → inapplicability
• Obstacles for transparency
Discussion: Predictive policing in the big data era
Chair: prof. dr. Wim Hardyns

Discussion

Technology is value-free (neutral)

The question is not whether or not we can predict the future. The question is more fundamental: what do we want to do with this knowledge?
• Based on linking data and data fusion, we can provide you a social credibility score, but do we want this?
• Based on new and emerging data collection methods, we can fine any non-paying parked car, but do we want this? And we can fine every car that drives too fast, but do we want this?

Discussion

Technology is the means, it is not the end goal

"There is no soul in the new machine" (Corbett & Marx, 1991)
Discussion

Methodology is fallible

- There are some things big data can never quantify, but
- there are also quantities that human can never phathom