

Inhaltsverzeichnis

1	Einleitung	3
2	Grundlagen und Notwendigkeit	4
2.1	Zahlen aus Kriminalstatistiken	4
2.2	Mögliche Problemfelder der Kriminologie	4
2.3	Predictive Policing	4
2.4	Googles Tensorflow	5
2.4.1	Integration in Python	5
2.5	Forschungslücke	5
3	Maschinelle Lernmethoden zum Forecasting	6
3.1	Altbewährte Methoden	6
3.1.1	Crime Mapping	6
3.2	Neuartige Methoden des maschinellen Lernens	6
3.2.1	Random Forest	6
3.2.2	Konvolutionales Neuronales Netzwerk	6
3.2.3	Weitere maschinelle Lernmethoden	6
3.3	Linearität und Nichtlinearität	7
3.4	Behandlung der Forecast-Errors	7
4	Empirie/ quantitative Auswertungen	8
4.1	Datensets	8
4.2	Auswertungen	8
4.3	Unterschiedliche Resultate in Datensets	8
5	Diskussion, Limitationen und Empfehlungen	9
5.1	Evaluation und Relevanz	9
5.2	Umsetzung in die Praxis	9
5.3	Politik und Recht	9
5.3.1	Datenbeschaffung	9
5.3.2	Zivil- bzw. Privatsrechte	9
5.3.3	Deutschlands Rolle in der Digitalisierung	9
5.4	Ausblick	10
6	Fazit	11
7	Anhang	11
8	Ehrenwörtliche Erklärung der Abschlussarbeit	11

Zusammenfassung

In der folgenden Masterarbeit dreht es sich um die Frage, inwiefern sich die Digitalisierung allgemein auf konventielle Berufe und im Speziellen auf den Beruf des Kriminalanalysten auswirkt. Die im Zuge der Digitalisierung häufig angewandten maschinellen Lernmethoden besitzen dabei das Potential, riesige Datenmengen in einer angemessenen Zeit nach Mustern zu durchsuchen und dabei Hinweise zu geben, wie einer Problematik entgegengewirkt werden kann. So könnten diese Analysen auch bei der Ergreifung von Verbrechern behilflich sein (*Predictive Policing*).

Bereits seit den 20er-Jahren wird versucht, die Wahrscheinlichkeit der Rückfälligkeit von Tätern bei möglichen Bewährungsstrafen miteinzubeziehen und diese dadurch zu klassifizieren. Auch bei der Ermittlung eines angemessenen Strafmaßes wird diese Wahrscheinlichkeit in Betracht gezogen. Die daraus resultierende Hauptfrage der folgenden Arbeit wird es sein, inwiefern sich alte Methoden (bspw. *Crime Mapping*) durch neuartige (bspw. *Random Forests/Neuronale Netzwerke*) ergänzen bzw. gänzlich ersetzen lassen, wie dies in der Realität evaluiert werden sollte und ob bzw. an welchen Positionen der Mensch im Zuge dessen überhaupt noch als Arbeitskraft benötigt wird.

Auch um die Frage einer angemessenen Veranschaulichung der Methodik wird es in der folgenden Arbeit gehen, da die Resultate für Entscheidungsträger so aufschlussreich wie möglich dargestellt werden sollten, um diesen sowohl die Legitimation ihrer Entscheidung als auch die Evaluation dieser zu vereinfachen. Dabei wird es zu Limitationen kommen, da sich durch maschinelle Lernmethoden häufig Umstände ergeben, die ein genaues Erfassen der Vorgehensweise dieser Methoden erschweren und so Rückschlüsse auf einzelne Variablen nur bedingt möglich sein werden, sodass die Vor- und Nachteile vor dem Hintergrund dieser Intransparenz kritisch aufgewogen werden müssen. Eine große Herausforderung ist es nämlich, Skeptiker von dem Potential zu überzeugen, die diese Methoden mit sich bringen.

Wichtige Angelegenheiten wie bspw. der Datenschutz und die damit einhergehende Wahrung der Privatsphäre müssen zusätzlich diskutiert werden, da es den Menschen Unbehagen bereitet, wenn mit ihren Daten nicht sorgfältig und anonym umgegangen wird. Die daraus resultierende Frage, was dabei rechtens ist und wie sich Gesetze im Bereich der Digitalisierung anpassen bzw. entwickeln sollten, um die perfekte Balance zwischen Persönlichkeitsrechten und öffentlicher Sicherheit zu finden, wird somit vom Autor berücksichtigt. So soll letztendlich mit dieser Arbeit dazu beigetragen werden, den Polizeiämtern weltweit eine Hilfestellung bei der Allokation begrenzter Ressourcen zu geben. Denn Straftaten zu verhindern ist in jeglicher Hinsicht kostengünstiger, als sie im Nachhinein aufzuklären.

1 Einleitung

- persönliche Motivation
- Vorhersagen sind in der Wirtschaft bereits stark vertreten (Verkauf/ Börse/ Trends in Mode und Musik/ etc.)
- ist dann auch menschliches Verhalten vorhersagbar oder zu zufällig/ irrational? (Perry, 3)
- > gibt es theoretische/ praktische Rechtfertigung/ Relevanz für diese Arbeit?
- auf welche Methoden wird schon lange gesetzt und wie lassen sich diese sinnvoll ergänzen? (von Heuristik zu modernen Algorithmen)
- wie lassen sich verschiedenste Daten/ Variablen miteinander in Verbindung bringen? (NIJ-Journal)
- wie ist der Status in Deutschland und weltweit -> Beispiele aufführen, wo es schon klappt(Random Gunfire in Richmond, Burglaries in Arlington (NIJ-Journal)?
- > viele Pilot-Projekte (BW, BY, BLN, Rhein-Ruhrgebiet mit PRECOPS und SPSS Modeller
- > siehe Wiki: Predictive Policing)
- Bezug zur Seismologie (UCLA)
- aktuell: Facebook-Skandal mit Cambridge Analytica
- 1984 von Orwell
- im Groben: siehe Zusammenfassung

2 Grundlagen und Notwendigkeit

2.1 Zahlen aus Kriminalstatistiken

- Interpol, Europol, BKA, BfV, etc. für die Notwendigkeit
- hochkomplexe und umfangreiche Datenmengen, insbesondere durch Vernetzung der Ämter für innere Sicherheit

2.2 Mögliche Problemfelder der Kriminologie

- welche können sowohl mit altbewährten als auch mit neuartigen Methoden angegangen werden (Perry, xv):
 - Identify areas at increased risk, Identify geographic features that increase the risk of crime, Find a high risk of a violent outbreak between criminal group, Identify individuals who may become offenders, Identify suspects using a victim's criminal history or other partial data (e.g., plate number)
 - is it a serial perpetrator? are there important persons or anchor places in the killers milieu? → automatisiertes Erkennen durch vernetzte Datenbanken

2.3 Predictive Policing

(aus Perry, S. xiv):

Folgende Anwendungsgebiete bestehen seit Langem, Versuch durch maschinelles Lernen und vernetzte Datenbanken auf große Datensätze zu übertragen (Casady, NIJ-Journal):

- Methods for predicting crime, Methods for predicting offenders, Methods for predicting perpetrators' identities, Methods for predicting victims of crimes
- Möglichkeiten, Erkenntnisse in Praxis umzusetzen? (bspw. mehr Polizeieinsatz an high-risk Orten (allgemein) oder andersartigen Polizeieinsatz (mehr spezifisch, bspw. anders/-stärker bewaffnet) (Perry))
- führende Software PredPol vorstellen, auch PreCobs, ESRI's ArcGIS, SPSS Modeller OpenSource-Software wie CrimeStat III
- weitere Autoren/ Modelle:
 - Berk (2013, Berechnung eines Rückfalls)
 - Chan (2015, Werden Social Scientists von Big Data Analysten abgelöst?)
 - Jeff Brantingham, Zusammenarbeit zwischen UCLA und LAPD
 - Craig D. Uchida

2.4 Googles Tensorflow

zum Aufbau von Modellen aus dem Bereich der maschinellen Lernmethoden

2.4.1 Integration in Python

das Programm Python vorstellen (Version, auch Rechnerkapazitäten etc.)

2.5 Forschungslücke

- Inwiefern lassen sich alte Methoden durch neuartige ergänzen bzw. gänzlich ersetzen?
- > Wird der Mensch überhaupt noch als Arbeitskraft benötigt im Bereich Kriminalanalyse?

3 Maschinelle Lernmethoden zum Forecasting

3.1 Altbewährte Methoden

3.1.1 Crime Mapping

- spatiale Analysen
- Hot-Spot Analysen
- Near-Repeat-Effekte
- Nachteil: zeigen nur vergangene Verbrechen, das Predicting selbst bleibt Betrachter selbst überlassen

3.2 Neuartige Methoden des maschinellen Lernens

3.2.1 Random Forest

- kein Overfit/ adaptiv/ Rückschlüsse auf einzelne Variablen durch den 'Increase in Forecasting Error/ Partial Dependence Plot' (Weglassen der Variable) möglich (Berk, 2013)
- Vergleich zu ähnlicher, früherer Methode

3.2.2 Konvolutionales Neuronales Netzwerk

- Overfit möglich
- > bereits erste Versuche im Bereich Natural Language Processing mit Wort-Vektoren durchgeführt:
Input: Fallbeschreibung
Output: Kategorie des Vorfalls
Hier den Vergleich zu früher ziehen, wo Kategorien u.U. händisch eingetragen werden mussten (Quellen?)

3.2.3 Weitere maschinelle Lernmethoden

- Social Network Analysis, Stochastic gradient boosting (Overfit/ adaptiv), Logistische Regression (nicht adaptiv, nur zwei Outputklassen)
- wichtig, dass nicht Apples mit Oranges verglichen werden (siehe Tuning-Parameter und andere Charakteristiken der einzelnen Methoden wie bspw. gute Performance bei lediglich kleinen/ großen Datensätzen)
- > muss vergleichbar sein
- allgemein: siehe Hastie, Perry

3.3 Linearität und Nichtlinearität

- je mehr Nichtlinearität, desto genauer, jedoch Gefahr des Overfittings

3.4 Behandlung der Forecast-Errors

- sollten diese gleich behandelt werden bei Klassifikationen?

-> Betrachtung der Auswirkung einer Fehleinschätzung (Kosten eines Toten durch Mord schwer vergleichbar/ messbar)

-> wie könnte man Fehler unterschiedlich gewichten? Bestimmung der Cost Ratio (False Negatives versus False Positives, Kosten eines unvorhergesagten Mordes wiegen schwerer als Kosten eines 'zu lange Verurteilten')

4 Empirie/ quantitative Auswertungen

4.1 Datensets

- Kriminaldatensätze aus Kaggle (San Francisco (878050x14; 2003-2015), Chicago (>6Mx17; 2001-2017), Baltimore (276530x14; 2012-2017), Seattle (1.5Mx18; 2010-2018))
- > welche Variablen sind typischerweise vorhanden (Zeit/ Ort/ Code/ Beschreibung/ Waffe)? Sind diese relevant? Wie ist die Datenqualität (Reliabilität, Validität, Objektivität)? Fehlen viele (data censoring), sind sie systematisch verzerrt (systematic bias)?

4.2 Auswertungen

- Confusion Table bei Klassifikationsproblemen

4.3 Unterschiedliche Resultate in Datensets

- die drei verschiedenen Datensätze auf Unterschiede vergleichen und diese versuchen zu erklären, falls vorhanden (evtl. auch Erkenntnisse zusammenführen? Data Fusion aus den unterschiedlichsten Variablen?)
- typische Charakteristiken der Städte wie Einwohner/Polizisten-Verhältnis
- falls möglich, Datensatz zu LA, da Vorreiterstellung

5 Diskussion, Limitationen und Empfehlungen

5.1 Evaluation und Relevanz

- anhand welcher Kennzahlen sollten die Methoden evaluiert werden
 - Crime Rates gesunken/ Arrest Rates gestiegen/ sozialer Einfluss, bspw. Angstempfinden der Gesellschaft gesunken
- sind Ergebnisse durch Empirie relevant (in absoluten Zahlen), besitzen also einen tak-tischen Nutzen oder wurde lediglich die Performance hochgeschraubt ohne wirklichen Mehrwert? Saisonale Schwankungen?
 - Einschätzung, ob Einführung neuartiger Praktiken deutlichen Zuwachs mit sich bringen ggü. altbewährten Praktiken

5.2 Umsetzung in die Praxis

- Empfehlungen für Stakeholder (Käufer, Verkäufer, Entwickler)
 - wie sollen Erkenntnisse aus der Empirie umgesetzt werden in die Praxis?
 - Aufpassen, dass es nicht zu sehr in die Richtung 'Minority Report' abdriftet, wo Menschen für Verbrechen verklagt werden, die sie (noch) nicht begangen haben
- vom Back-End zum Frontend: möglicherweise über Ruby (Live-Mapping/ Updates auf mobile Geräte)
- Schulungen für Anwender

5.3 Politik und Recht

5.3.1 Datenbeschaffung

- auf was sollte stets bei der Datenbeschaffung geachtet werden?
- aufpassen: Facebook-Skandal mit Cambridge Analytica
- 1984, George Orwell
- Transparenz der Vorgänge äußerst wichtig (NIJ-Journal)

5.3.2 Zivil- bzw. Privatsrechte

- inwiefern sind Zivil- bzw. Privatsrechte bei den unterschiedlichen Methoden betroffen, v.a. bei Social Network Analysis?
- auf fragliche Paragraphen eingehen? oder zu rechtlich?

5.3.3 Deutschlands Rolle in der Digitalisierung

- auf neu geschaffenes 'Digitalministerium' eingehen (Dorothee Bär als neue 'Digitalministerin') sowie Deutschlands Rolle im Bereich Digitalisierung insgesamt

- > aufpassen, dass es nicht zu weitläufig wird
- inwieweit sollte die Gesellschaft insbesondere Skeptiker miteingebunden werden beim Aufbau von Software/ Gangart, um Zweifel auszuräumen (NIJ-Journal)

5.4 Ausblick

- mit Mythen aufräumen, die weit verbreitet sind was Predictive Policing angeht, wie Kristall-Kugel (Perry, xix)
- wie wichtig wird der Einsatz eines Menschen überhaupt noch sein? Empfehlungen zur besseren Zusammenarbeit zwischen Mensch und Maschine (Perry, xxiii)
- Wie ist der wissenschaftliche und praktische Beitrag zu bewerten?
- welche Anwendungsgebiete bestehen neben der Polizei? Sicherheitsdienste, Militär, Beratung und Software-Entwicklung

6 Fazit

Referenzen

7 Anhang

8 Ehrenwörtliche Erklärung der Abschlussarbeit