Crime Detection in Text for a Better Moral Reasoning System

Radoslaw Komuda[†], Michal Ptaszynski[‡], Rafal Rzepka[†], Kenji Araki[†]

[†] Graduate School of Information Science and Technology   [‡] Department of Computer Science
Hokkaido University   Kitami Institute of Technology
[†] {komuda,rzepka,araki}@ist.hokudai.ac.jp,   [‡] ptaszynski@cs.kitami-it.ac.jp

Abstract

We begin this paper with highlighting the importance of Artificially Intelligent in supporting human efforts in mining web resources for crime-related conversation. By introducing Lawrence Kohlberg's theory of stages of moral development, we identify this work as preliminary research and a part of an ongoing research. Thirdly, we discuss our experiences in extracting drug- and traffic-related offences from the British law system and ranking them in a weighted model. Finally, we describe our methods to broaden system's results and present sample query result on British National Corpus.

Keywords — Crime Detection, British Law System,

1. Introduction

Machines have been used as tools to support human efforts in many fields like industry, information gathering and processing. Recently highly specialized machines, such as computers and robots are also taking over human roles in many parts of these fields. The development of Internet resulted in new challenges, especially for law enforcement, such as online frauds or identity theft. With their computational capabilities, computers are able to find patterns difficult to discover for humans. However, social network services, blogs, dark web fora – all of these virtual places contain conversations containing potential crime references. Therefore, detecting them, recognizing and evaluating is an important endeavor.

This paper introduces preliminary research on a system designed to automatically detect such crime-related references in text.

2. Kohlbergian agent idea introduction

Lawrence Kohlberg was an American psychologist and the author of a theory of scale of moral development [1]. Known as the "moral equivalent of Piaget's theory of cognitive development", Kohlberg's theory states that there are 6 consecutive and rare to regress levels of moral reasoning:

- stage 1: obedience and punishment driven morality in which the worse the punishment for the act, the more "wrong" an action is perceived to be,
- stage 2: self-interest driven approach searching for the direct benefit from an action,
- stage 3: "good" determined by social consensus on approval or disapproval to certain actions,
- stage 4: law and social order driven morality that ensures maintenance of a functioning society,
- stage 5: laws and rules considered a "social contract" and though this approach respects individuality – the majority decides,
- stage 6: universal ethical principles and abstract moral reasoning level.

Using Kohlberg's theory as a path to follow [2], this paper introduces stage 4-based module development.

3. Crimes and penalties catalogue creation

For this research we use Welcome to Gov.UK website [5]. We have decided to manually extract drugs- and traffic-related felonies, infractions, violations and / or misdemeanors from the English (England and Wales) law system.

For drugs-related penalties, we were able to collected 60 initial entries, covering possession and supply and production of drugs of different classes, incl. cocaine and LSD, cannabis and ketamine or khat and anabolic steroids. Traffic-related crimes catalogue cover 20 entries, from some of the most serious, such as "causing death by dangerous driving", "failing to stop after an accident," "failing to report an accident", or "driving without insurance", through less socially harmful, such as "speeding" or "cycling on the pavement."

4. Ranking creation and penalty weighting

The maximum penalties for drug possession,

supply (dealing) and production in the England and Wales' law system depends on what type or 'class' the drug is, having them grouped A through C (plus additional "temporary class") depending on how strong they are differing from lifetime sentence to a mere confiscation.

For traffic-related offences, the penalties start from a plain 500 GBP fine and go all the way up to 14 years of imprisonment.

Second step with working with the created catalogue was to rank actions accordingly with these penalties. For the purpose of this ranking creation we have decided to normalize our findings:

1. prison penalties were normalized by dividing each value by the maximum penalty expected; lifetime sentence was counted as 100 years,

2. fines were calculated with the above manner but unlimited fine was determined as 10,000 GBP, because this amount reflected its nature,

3. driving disqualification was ranked with 1 point for obligatory disqualification, 0.5 point for discretionary and 0 for lack of disqualification penalty,

4. only maximum value of the penalty points (pp) was taken for consideration, i.e. "3-11 pp" wound be ranked higher than "5-10 pp" and where the penalty was either 3-6 pp or a fixed penalty of 3 pp – the ranking would position the offence among other 6-pp-valued ones.

After penalty normalization we have created a weighted model of the ranking, multiplying normalized values by 100, 7, 2 and 1 for prison, fine, disqualification and pp penalties respectively. This approach resulted in a unified ranking, making it possible to compare offences from not only the same crime group, but also – between different offence types.

5. Weighted model result analysis

In the final quantification (Appendix 1) – being a summation of the weighted values – A-class drug supply and production with its lifetime sentence and unlimited fine was the top offence in the ranking. Next five offences, namely causing death by dangerous driving, causing death by careless driving under the influence of drink or drugs, B-class drug supply and production, C-class drug supply and production and temporary-class drug supply and production shared the same score for prison (up to 14 years) and unlimited fine penalties but the weighted model has put driving-related events higher in the ranking. Though the potential harm and reach for drug-reletad offences might be considered higher, the ranking's basic goal is to assess consequences for the Actor and at current stage of the research it does not take social harm into consideration. This situation repeated itself in later part of the ranking, putting dangerous driving before C-class drug possession offence.

6. Common English lexicon creation

Our initial experiments with the law-language inputs within the British National Corpus has proven to limit the number of entries extracted, e.g. "failing to stop after an accident" returns only 6 entries, while "hit and run" query results in 28 entries being returned. "Cannabis" on the other hand returns 400 results but "marijuana" expands the list of entries by over 33%. Because of these findings, we have decided to expand our offence catalogue with WordNet [3] and ConceptNet [4] to find words and phrases connected with the initial law terminology.

7. Different search methods used

We performed our crime detection experiment with a keyword search with the crimes names across the corpus, which resulted in high precision but rather low recall. To enhance these results, we have enriched the queries and used not only the crime itself, but also – included details, i.e. searched not for "A-class drug possession" but named every drug in that group and did a separate search for each one of them, which increased the recall. After excluding English stop words, system's precision was increased even more. Getting our list of crime synonyms extended with the ConceptNet and WordNet help, we improved the results even further.

Table 1 and Table 2 present sample frequency results for different queries performed on the British National Corpus:

Table 1 Sample results for drug-related queries

Query:	Frequency:
cocaine	634
ecstasy	592
heroin	923
LSD	165
magic mushroom	1
methadone	77
methamphetamine	1
amphetamine	131
cannabis	418
ketamine	59

Table 2 Sample results for traffic-related queries

Query:	Frequency:
dangerous driving	63
careless driving	43
inconsiderate driving	1
drink driving	
driving while disqualified	1
driving without insurance	77
speeding	1
seat belt offence	131

5. Conclusions and future work

This paper introduced a system for crime recognition in text. By using a weighted model approach, we were able to compare actions from different categories in an independent manner, with the usage of commonly respected factors. Our research proved that when working with law documents alone one can get a limited amount of data and by enriching law-language entries with common English words and phrases (i.e. synonyms) broader insight into the corpus' content is provided.

In future work we wish to develop both systems' weighted offense ranking and the lexicon database. We consider testing systems capabilities on other British corpora as well, i.e. British Blog Corpus.

References

[1] Lawrence Kohlberg, (1981) „Essays on Moral Development", Vol. I The Philosophy of Moral Development. San Francisco, CA: Harper & Row. (1981).
[2] Radoslaw Komuda, (2010) „Kohlberg's Theory of Moral Development Stages as a Path to Follow In Machine Ethics Research", Language Acquisition and Understanding Research Group (LAU) Technical Reports, pp. 6-10.
[3] WordNet, https://wordnet.princeton.edu
[4] Concept Net, http://conceptnet5.media.mit.edu
[5] Welcome to Gov.UK, https://www.gov.uk

Appendix 1 Weighted model calculation and final quantification:

offence:	penalty:				weighted model:				quantification:
	prison:	fine:	disqual.:	pp:	prison:	fine:	disqual.:	pp:	
A-class drug supply and production	100	10,000	--	--	7.14	4			746.29
causing death by dangerous driving	14	10,000	1	11	1.00	4	1	1	135.00
causing death by careless driving under the influence of drink or drugs	14	10,000	1	11	1.00	4	1	1.00	135.00
B-class drug supply and production	14	10,000	--	--	1.00	4			132.00
C-class drug supply and production	14	10,000	--	--	1.00	4			132.00
temporary-class drug supply and production	14	10,000	--	--	1.00	4			132.00
A-class drug possession	7	10,000	--	--	0.50	4			82.00
B-class drug possession	5	10,000	--	--	0.36	4			67.71
dangerous driving	2	10,000	1	11	0.14	4	1	1.00	49.29
C-class drug possession	2	10,000	--	--	0.14	4			46.29
driving while unfit or failing to provide specimen for analysis	0.5	10,000	1	11	0.04	4	1	1.00	38.57
failing to stop after an accident or failing to report an accident	0.5	10,000	0.5	10	0.04	4	0.5	0.91	37.48

offence:	penalty:				normalization:				quantification
	prison:	fine:	disqual.:	pp:	prison:	fine:	disqual.:	pp:	
driving while disqualified	0.5	10,000	0.5	6	0.04	4	0.5	0.55	37.12
driving after refusal or revocation of licence on medical grounds	0.5	10,000	0.5	6	0.04	4	0.5	0.55	37.12
using a vehicle in a dangerous condition	0	10,000	1	3	-	4	1	0.27	34.27
careless and inconsiderate driving	0	10,000	0.5	9	-	4	0.5	0.82	33.82
driving without insurance	0	10,000	0.5	8	-	4	0.5	0.73	33.73
speeding	0	2,500	0.5	6	-	1	0.5	0.55	9.55
failure to have proper control of vehicle or full view of the road and traffic ahead, or using a hand-held mobile phone when driving	0	2,500	0.5	3	-	1	0.5	0.27	9.27
dangerous cycling	0	2,500	0	0	-	1	0	-	8.00
driving otherwise than in accordance with a license	0	1,000	0.5	6	-	0.4	0.5	0.55	4.75
traffic light offences	0	1,000	0.5	3	-	0.4	0.5	0.27	4.47
failing to identify driver of vehicle	0	1,000	0.5	0	-	0.4	0.5	-	4.20
no MOT certificate	0	1,000	0	0	-	0.4	0	-	3.20
careless cycling	0	1,000	0	0	-	0.4	0	-	3.20

offense:	penalty:				normalization:				quantification:
	prison:	fine:	disqual.:	pp:	prison:	fine:	disqual.:	pp:	
seat belt offences	0	500	0	0	-	0.2	0	-	1.60
cycling on pavement	0	500	0	0	-	0.2	0	-	1.60
temporary-class drug possession	0	0	--	--	-	0			-