Modelling Online Anomalous Behaviour using Big Social Data

Giles Oatley and **Tom Crick**
Department of Computing & Information Systems
Cardiff Metropolitan University
Cardiff, UK

Abstract

The rise of Web 2.0 and the popularity of social networking has facilitated the publishing of user-generated content on an exponential scale; its analysis is becoming increasingly important (and applicable) to the empirical study of society (and thus societal change). However, this "big social data" from social media platforms differs significantly from more traditional/formal sources. In this paper we focus our attention on modelling anomalous behaviour using big social data; for example, so-called 'unnatural' language with its poor language construction but also context dependent acronyms, jargon, "leetspeak" and profanity. We also include a discussion on language use related to pornography, as well as the criminal element of identity theft, fraud and deliberate deception.

Our long-term research goal is the development of complex (and adaptive) behavioural modelling and profiling using a multitude of online datasets; in this paper we highlight relevant tools for analysing these big dataset, and also present some novel ideas around a scalable solution to analysing swear words.

Introduction

With recent privacy and national security incidents – such as with Chelsea Manning, Wikileaks (The Guardian, Leigh, and Harding 2013) and Edward Snowden (Greenwald 2014) – bringing significant attention to profiling insider threats, there are now large-scale research efforts in developing new and robust techniques for modelling online behaviour and identity. There are numerous domains in which it is essential to obtain knowledge about user profiles or models of software applications, including intelligent agents, adaptive systems, intelligent tutoring systems, recommender systems, e-commerce applications and knowledge management systems (Schiaffino and Amandi 2009); we discuss our ideas relating to profiling complex behaviours elsewhere (Oatley and Crick 2014).

Big datasets from social networking platforms are now being used for a multitude of purposes, alongside the obvious advertising, marketing and revenue generation; increasingly for government monitoring of citizens[1,2,3], along with covert security, intelligence community and military user profiling. However, the publishing of user-generated content on an exponential scale has significantly changed qualitative and quantitative social research, with its analysis becoming increasingly important to the empirical study of society. There are interesting sociological uses of studying or mining big social data, for instance exploring cyber-physical crowds using location-tagged social networks or the study of personality with large-scale benchmark social datasets and corpora.

Big data from social networking sources differs significantly from more traditional sources. With the advent of the social web, for instance social networks, blogs, gaming, shopping and review sites, there are now orders of magnitude more data available relating to uncensored natural language, requiring the development of new techniques that can meaningful analyse it. This uncensored language is rich in 'unnatural' language (as opposed to 'natural' language, used in formal/traditional published media such as books and newspapers), defined as "*informal expressions, variations, spelling errors...irregular proper nouns, emoticons, unknown words*" (ULP 2011); it is also rich in context-dependent acronyms, jargon, "leetspeak" and swear words. Leet, also known as eleet or leetspeak, is an alternative alphabet for the English language that is used primarily on the Internet and in geek/cyber communities. It uses various combinations of ASCII characters to replace Latin script. For example, leet spellings of the word "leet" include *1337* and *l33t*; eleet may be spelled *31337* or *3l33t*. See Perea, Dunabeitia, and Carreiras (2008) for an discussion of leet from a cognitive processing perspective.

Representing Complex Behaviour and Personality

Advances in psychology research have suggested it is possible for personality to be determined from digital data (Pennebaker and King 1999; Vazire and Gosling 2004; Iacobelli

[1]Twitter Transparency Report 2014:
https://transparency.twitter.com/

[2]Facebook Global Government Requests Report 2014:
https://govtrequests.facebook.com/

[3]Google Transparency Report 2014:
http://www.google.co.uk/transparencyreport/

et al. 2011). Recent studies (Woodworth et al. 2012) have suggested certain keywords and phrases can signal underlying tendencies and that this can form the basis of identifying certain aspects of personality. Extrapolation suggests that by investigation of an individual's online comments it may be possible to identify individual's personality traits. Initial evidence in support of this hypothesis was demonstrated in 2012 by analysis of Twitter data for indicators of psychotic behaviour (Sumner et al. 2012). While in the past this has mainly been the textual information contained in blogs, status posts and photo comments (Blamey, Crick, and Oatley 2012; 2013), there is also a wealth of information in the other ways of interacting with online artefacts. For instance, it is possible to observe the ordering/timings of button clicks of a user. Several researchers have looked at personality prediction (e.g. Five Factor personality traits) based on information in a user's Facebook profile (Back et al. 2010; Golbeck, Robles, and Turner 2011) and speech (Chung and Pennebaker 2007; Tausczik and Pennebaker 2010), as well as also demonstrating significant correlations with fine affect (emotion) categories such as that of excitement, guilt, yearning, and admiration (Mohammad and Kiritchenko 2013). There are also several strands of related work based on the benchmark myPersonality Project[4] dataset (Celli et al. 2013), providing a platform for well-needed comparative studies.

Mairesse et al. (2007) highlighted the use of features from the psycholinguistic databases LIWC (Pennebaker, Francis, and Booth 2001) and MRC (Wilson 1988) to create a range of statistical models for each of the Five Factor personality traits (Norman 1963; Peabody and Goldberg 1989).

In previous work (Oatley and Crick 2014) we utilised these methods to develop a complex behavioural profile that included 'two faces' to model that we can have several different modes of operation (ego states). We performed our Five Factor analysis, and elaborated two sets of Five Factor results for each user. We chose Chernoff faces (Chernoff 1973) for the visual representation. The Five Factors are displayed as five features on a stylised face, where:

- Width of hair represents *Conscientiousness*;
- Width of eyes represents *Agreeableness*;
- Width of nose represents *Openness to experience*;
- Width of mouth represents *Emotional stability*;
- Height of face represents *Extraversion*.

Figure 1: Two faces for the same person. Each face represents a particular personality profile based on the Five Factors model

[4] http://mypersonality.org/

It should be noted that while researchers have continued to work with the Five Factors model, there are well known limitations (Eysenck 1992; Paunonen and Jackson 2000; Block 2010) that are often overlooked by researchers. In particular, it has been criticised for its limited scope, methodology and the absence of an underlying theory. However, attempts to replicate the Big Five in other countries with local dictionaries have succeeded in some instances but not comprehensively (Szirmák and De Raad 1994; De Fruyt et al. 2004). While Costa and McCrae (1992) claim that their Five Factors model "represents basic dimensions of personality", psychologists have identified important trait models, for instance Cattell's 16 Personality Factors (Cattell 1946) and Eysenck's biologically-based theory (Eysenck 1947).

Readability and General Measurements of English Text

There are several well-known measures of a text's readability, for instance the Gunning Fog index (Gunning 1968) and the Flesch-Kincaid tests (Kincaid, Fishburne, and Chissom 1975). These can also be used to profile an individual.

The Gunning Fog index indicates the number of years of formal education a reader of average intelligence would need to understand the text on the first reading and computes a scale (unreadable/difficult/ideal/acceptable/childish) based on the ratios of the counts of words to sentences and counts of complex words to words. The Flesch-Kincaid tests consider word length (numbers of syllables) and sentence length (numbers of words).

Fraud, Deception, Identity Theft and Bad Behaviour

Automatically detecting a fraudulent account in a software system is a complex and challenging task, requiring that an algorithm is able to grasp the meaning of a persons' social data without significant human intervention. Consider the complexities of detecting specific events, for instance marriage, anniversaries, payday, and so on, which are difficult features to represent and compute. This would be necessary to start to 'understand' what was happening or being represented in someone's digital footprint. Perhaps prior to a marriage there will be a list of potential words used in their social data, and a different (but complementary) set of words used afterwards, and we can thus triangulate on our best guess (obviously here we are ignoring direct knowledge of the date of the event, for instance through calendars, or time-stamped photos or albums tagged with 'wedding' and so on). To facilitate this, a significant amount of knowledge will need to be represented and organised, for instance in specific knowledge modelling or bespoke parsers and lexicons.

There is no doubt that profiling is difficult, and in relation to detecting (Five Factors) personality, Mairesse (2013) recommends learning to map user inputs to elements that can improve its interaction directly, that is, a sharply focused mode, and also that there should just be a standard machine learning pipeline, and not a 'personality-based' user model. Researchers from the knowledge representation/engineering

community will recognise these problems – the bottleneck of acquiring and appropriately representing domain knowledge – and this perhaps received its most practical explanation in Richter's (2003) concept of knowledge containers in case-based reasoning systems.

Consider the conclusions of the *Uncovering plagiarism, authorship and social software misuse* (PAN) stream of CLEF 2012[5] on detection of sexual predators within chat forums (Inches and Crestani 2012). *Porn Predators* was an evaluation lab (Sexual Predator Identification Competition) within a broader annual conference covering plagiarism, authorship and social software misuse. In relation to profiling technologies the conference covers author identification and verification and author profiling, including methods to answer the question whether two given documents have the same author or not, and also predicting an author's demographics. The tasks within *Porn Predators* were to:

(i) identify the predators among all the users conversations;

(ii) identify the specific part of the conversations most distinctive of predator behaviour.

In summary, for the first problem lexical and behavioural features should be used, although there is no unique method proposed, and for the second problem the most effective methods were those based on filtering on a dictionary or lexicon, partly due to the lack of ground truth for this specific problem.

The idea behind detecting fraud, or an insider threat, is that the profile of an individual changes in significant ways, and by developing a (knowledge) representational system rich enough we will be able to detect this. And so we currently use a range of psycholinguistic features, more complex behavioural or trait features, lying/deception, and linguistic styles, formality, readability and so on. Of course, there are simpler methods to detect bad behaviour related to fake accounts, for instance we might be interested in screening for non-UK cities, or the age of social networking accounts. While it is easy to automatically generate social networking profiles with followers and activity for a few dollars, and a range of links can be developed, certain features like the age of posts and photo uploads cannot be backdated to appear more established and credible.

The Language of Pornography

Colleagues carried out a research project investigating opinions on a range of topics related to pornography usage; a web-based questionnaire received over five thousand respondents (n=5490). Several of the questions were open-ended, for instance how the person became involved with the subject of pornography, their particular interests and so on, eliciting on occasion a number of detailed responses (c.2000 words). From the initial findings (Smith, Attwood, and Barker 2013), the data is ill-structured, with frequent usage of bad grammar and contains a large number of jargon (swear) words relating to pornography and sexuality.

[5]See: `http://www.clef-initiative.eu/edition/clef2012/working-notes`

An aim of the original study was the investigation of the usage of fantasy. This resonated with our general interest in determining behaviour from data, and so explored the language characteristics of the answers related specifically to fantasy. We analysed the respondents text using the psycholinguistic databases LIWC and MRC. The Dictionary of Affect in Language (DAL) (Sweeney and Whissell 1984) was also planned to be used, due to its specific uses for imagery-based language. We also used methods derived from LIWC and MRC to determine personality traits and measures such as formality and deception. We also wanted to get a general feel for the level of the text, and to also see if there were any correlations between literacy and readability.

Initially we focused on the specific questions that might reveal something about the role of fantasy. For instance, among the many options for the question "*What are your reasons for looking at pornography?*", among the list were the following:

(A) "*To see things I might do*";

(B) "*To see things I can't do*";

(C) "*To see things I wouldn't do*";

(D) "*To see things I shouldn't do*".

The '*can't*' and '*wouldn't*' choices clearly indicate respondents utilising pornography more strongly as a form of fantasy. For this we explored the Five Factors personality traits, in particular expecting some correlation with the *Openness to Experience* factor.

	A	B	C	D
A	1			
B	-0.72974	1		
C	-0.46635	-0.06469	1	
D	-0.33821	0.08321	0.091183	1

Table 1: Correlation between question items (where: A="*To see things I might do*"; B="*To see things I can't do*"; C= "*To see things I wouldn't do*" D="*To see things I shouldn't do*")

Analysis is ongoing, with the results to be published in the near future; however there appears to be a strong negative correlation between participants who chose "*A. To see things I might do*" versus "*B. To see things I can't do*", as originally hypothesised. What was less convincing was our analysis of the Five Factors, and we put this down to the measures we used from Mairesse et al. (2007) being derived from a somewhat different corpus. We are currently concentrating on the lower level features from LIWC, MRC and DAL.

Among the categories of pornography for "*What kinds of sexually explicit materials do you access*" were the options *Fiction Sites*, *Sex Blogs* and *Stories*. Table 2 shows the correlation between these categories. All of these pornographic sites are fiction or story-based, and we examined the literary features of the respondents' replies to see if this group of respondents write in a more sophisticated way.

In all cases (see Figures 6–12), it appears that the users favouring the text/story-based pornography have higher values than their non-text contemporaries.

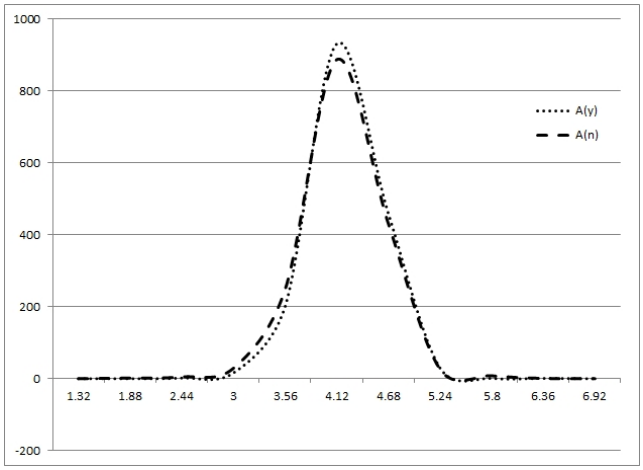

Figure 2: Openness to experience for A(y) (dotted) versus non-A (dashed)

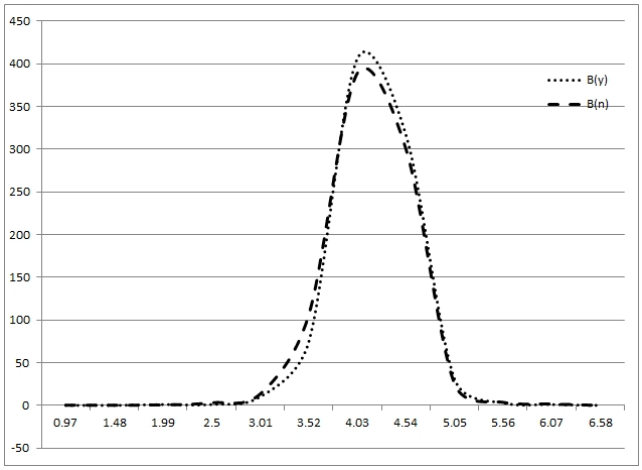

Figure 3: Openness to experience for B(y) (dotted) versus non-A (dashed)

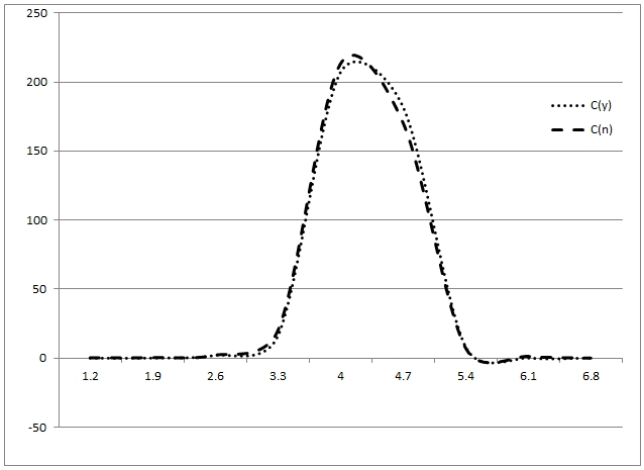

Figure 4: Openness to experience for C(y) (dotted) versus non-A (dashed)

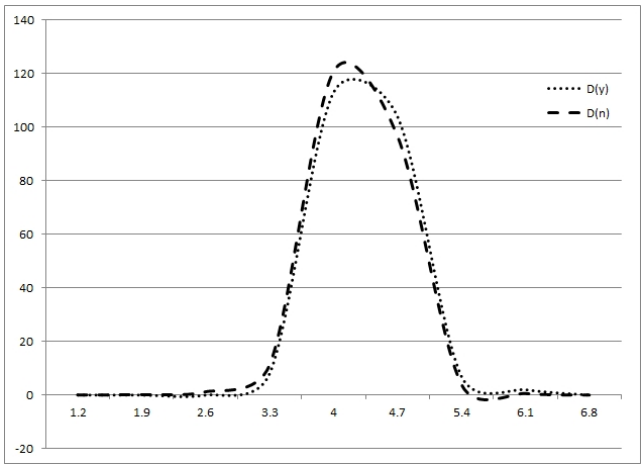

Figure 5: Openness to experience for D(y) (dotted) versus non-A (dashed)

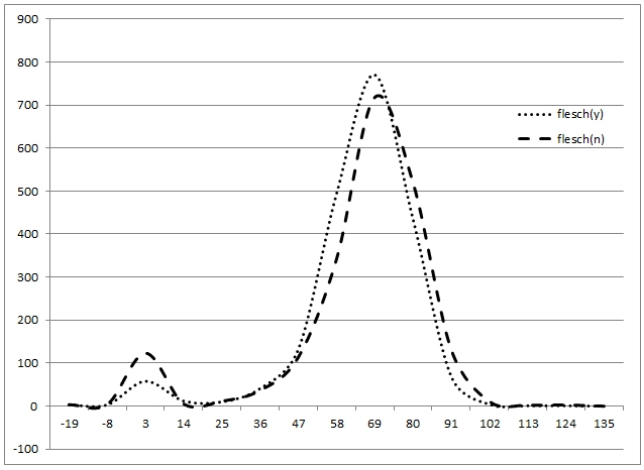

Figure 6: Flesch measure. Fiction readers (dotted) versus non-fiction readers (dashed)

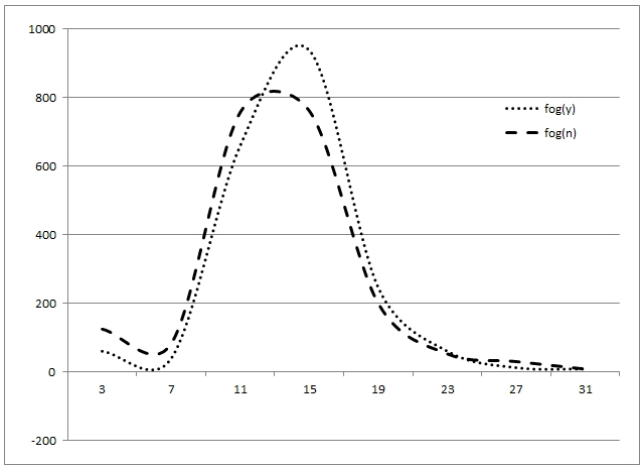

Figure 7: Fog index. Fiction readers (dotted) versus non-fiction readers (dashed)

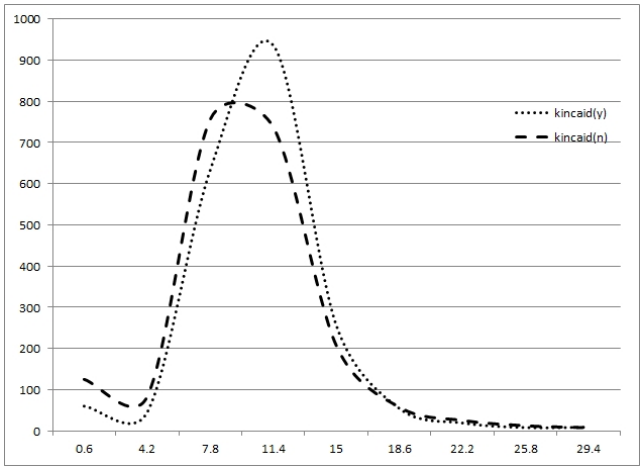

Figure 8: Kincaid measure. Fiction readers (dotted) versus non-fiction readers (dashed)

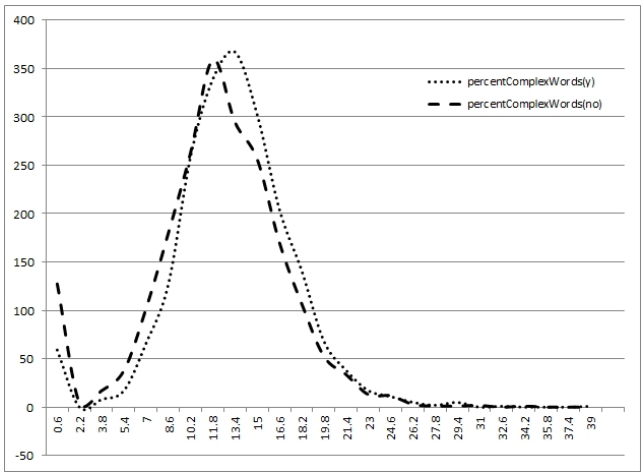

Figure 9: Percent complex words. Fiction readers (dotted) versus non-fiction readers (dashed)

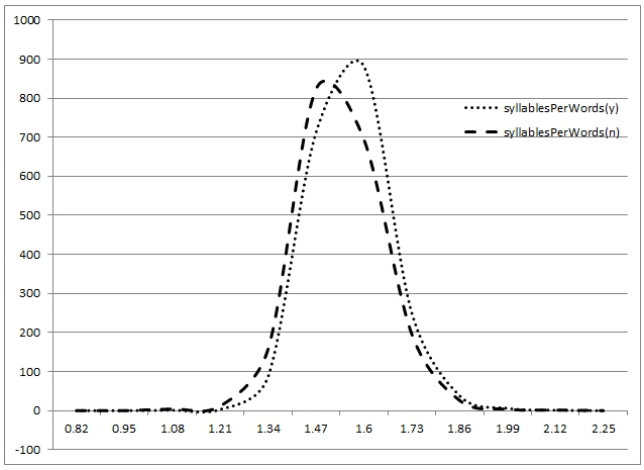

Figure 10: Syllables per word. Fiction readers (dotted) versus non-fiction readers (dashed)

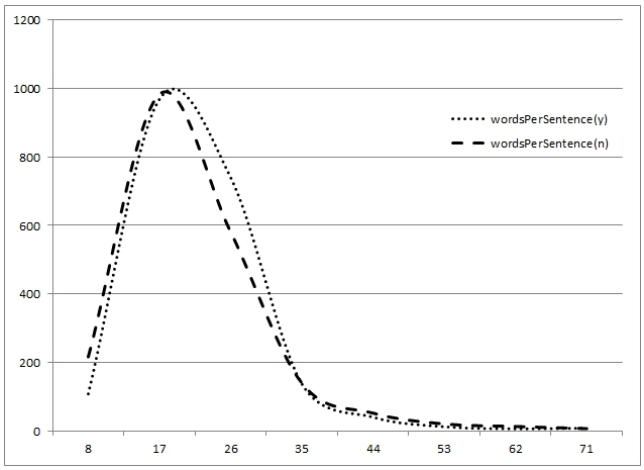

Figure 11: Words per sentence. Fiction readers (dotted) versus non-fiction readers (dashed)

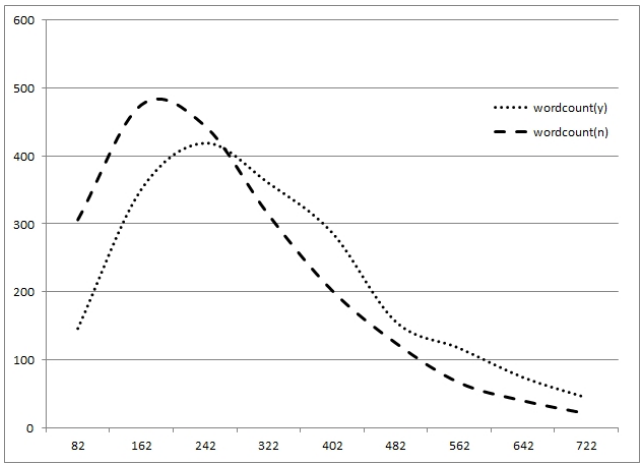

Figure 12: Word count. Fiction readers (dotted) versus non-fiction readers (dashed)

	Fiction	Blogs	Stories
Fiction	1		
Blogs	0.208086	1	
Stories	0.192054	0.041866	1

Table 2: Correlation between text-related options.

Disambiguating Profanity

WordNet[6] is a large lexical database of English; nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept, and each synset is interlinked by means of conceptual-semantic and lexical relations. Words that are found in close proximity to one another in the network are semantically disambiguated. WordNet Affect[7], a hierarchical set of emotional categories, and SentiWordNet[8], synsets are assigned sentiment scores (positivity, negativity, objectivity), are built on top of WordNet.

Millwood-Hargrave's (2000) study for Ofcom (formerly, the Broadcasting Standards Commission), the UK's regulatory and competition authority for the broadcasting, telecommunications and postal industries, was designed to test people's attitudes to swearing and offensive language, and to examine the degree to which context played a role in their reactions. Included in the report was attitudes towards swearing and offensive language 'in life', including a range of swear words and terms of abuse. Appendix 2's 'list of words' contained positions of the top swear words (categorised as "*very severe*", "*fairly severe*", "*quite mild*" and "*not swearing*") along with their ranking from 1998 to 2000.

The study of swear words has a longstanding position in linguistics, with the academic journal *Maledicta: The International Journal of Verbal Aggression* running from 1977 until 2005. Maledicta was dedicated to the study of the origin, etymology, meaning, use and influence of vulgar, obscene, aggressive, abusive and blasphemous language. Unfortunately we do not have resources such as databases in the literature; furthermore, WordNet does not contain the range of swear words we encountered in our data and is no use for disambiguating our text. Wikipedia, however, fared much better; but even better than these were Roger's Profanisaurus and Urban Dictionary.

Roger's Profanisaurus[9] is a lexicon of profane words and expressions; the 2005 version (the Profanisaurus Rex), contains over 8,000 words and phrases, with a further-expanded version released in 2007. Unlike a traditional dictionary or thesaurus, the content is enlivened by often pungent or politically incorrect observations and asides intended to provide further comic effect.

Urban Dictionary[10] is a Web-based dictionary that contains nearly eight million definitions as of December 2014. Originally, Urban Dictionary was intended as a peer-reviewed dictionary of slang or cultural words or phrases not typically found in standard dictionaries, with words or phrases on Urban Dictionary having multiple definitions, usage examples and tags.

We created different gazetteers related to rude words; one list was based on Wikipedia entries, and another on lists from Urban Dictionary. The Wikipedia list was created from link text on the Wikipedia porn sub-genre page[11] (link "anchor text" is a typical approach in semantic relatedness studies). This was comprised of 250 words. The Urban Dictionary list was created from the "sex" category[12] (by no means exhaustive – it is a fraction of the pornography-related terms in Urban Dictionary). This was comprised of 156 words. We implemented two metrics for rude words, the key idea of which is to have a simple mathematical model that enables us to estimate the life-history value of a token.

1. **IndexOfRudeWords:** For each sentence we have an ordered set {*rude1,rude2,rude3*}. This would be mapped to the following (rude words for sentences appearing sequentially); only the first mention is added to list.

```
{
    rude1 -> .25
    rude2 -> .5
    rude3 -> .75
}
```

   We could say that this represents the relative position in terms of rude word occurrence, and therefore we can call this a normalised rude word index.

2. **Sentence Index:** For each sentence in a block of text, the index of the sentence is used.

```
{
    // rude1 mentioned
    // in sentences 1 and 3.
    rude1 -> {1,3}
    // etc...
    rude2 -> {2}
}
```

For (1) and (2) we computed an average metric for each rude word over all sentences, additionally with the possibility of excluding those below some frequency (per word) and/or excluding responses where the number of rude words was very low. An alternative would be representing as a fractional token index.

[6] http://wordnet.princeton.edu/

[7] http://wndomains.fbk.eu/wnaffect.html

[8] http://sentiwordnet.isti.cnr.it/

[9] http://www.viz.co.uk/profanisaurus.html

[10] http://www.urbandictionary.com/

[11] http://en.wikipedia.org/wiki/List_of_pornographic_sub-genres

[12] http://www.urbandictionary.com/category/sex

Conclusions and Future Work

Existing NLP tools are known to struggle with unnatural language: "*demonstrated that existing tools for POS tagging, chunking and Named Entity Recognition perform quite poorly when applied to tweets*" (Ritter et al. 2011) and "*showed that [lengthening words] is a common phenomenon in Twitter*" (Brody and Diakopoulos 2011), presenting a problem for lexicon-based approaches. These investigations both employed some form of inexact word matching to overcome the difficulties of unnatural language. We have not yet used inexact string matching or made use of a leetspeak parser; this will form part of future work.

To assist with the ongoing knowledge modelling problem in this domain we recognise the need to utilise specific lexicons that keep pace with the language used, for instance the use of Urban Dictionary to resolve swear words. We need to study precisely how in what manner this resource keeps pace with popular culture.

There are numerous other lists of pornographic words, which we compiled from miscellaneous sources; however, we are mainly interested in sources such as Wikipedia and Urban Dictionary as these are maintained by a community that uses the words in social networking. In this way, we do not have to concern ourselves about this knowledge engineering process, merely concern ourselves about the representation and quality of meaning or definitions. We will in future work incorporate the voting scores available on Urban Dictionary, and look to fully integrate resources such as Roger's Profanisaurus.

Future work will continue with the range of techniques to analyse content for words/phrases associated with certain key emotions/states. We will further develop the bottom-up use of two and three word n-grams and top-down approach using the LIWC, MRC and DAL databases (Iacobelli et al. 2011).

References

Back, M. D.; Stopfer, J. M.; Vazire, S.; Gaddis, S.; Schmukle, S. C.; Egloff, B.; and Gosling, S. D. 2010. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science* 21(3):372–374.

Blamey, B.; Crick, T.; and Oatley, G. 2012. R U :-) or :-(? Character- vs. Word-Gram Feature Selection for Sentiment Classification of OSN Corpora. In *Research and Development in Intelligent Systems XXIX, Proceedings of the 32nd SGAI International Conference on Artificial Intelligence (AI-2012)*, 207–212. Springer.

Blamey, B.; Crick, T.; and Oatley, G. 2013. 'The First Day of Summer': Parsing Temporal Expressions with Distributed Semantics. In *Research and Development in Intelligent Systems XXX, Proceedings of the 33rd SGAI International Conference on Artificial Intelligence (AI-2013)*, 389–402.

Block, J. 2010. The Five-Factor Framing of Personality and Beyond: Some Ruminations. *Psychological Inquiry* 21(1):2–25.

Brody, S., and Diakopoulos, N. 2011. Cooooooooooooool-lllllllllllll!!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*, 562–570.

Cattell, R. B. 1946. *The description and measurement of personality*. Harcourt, Brace & World.

Celli, F.; Pianesi, F.; Stillwell, D.; and Kosinski, M., eds. 2013. *Workshop on Computational Personality Recognition (Shared Task)*.

Chernoff, H. 1973. The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of the American Statistical Association* 68(342):361–368.

Chung, C., and Pennebaker, J. W. 2007. *Social Communication: Frontiers of Social Psychology*. Psychology Press. chapter The psychological function of function words, 343–359.

Costa, P. T., and McCrae, R. R. 1992. *Neo PI-R Professional Manual*. Psychological Assessment Resources.

De Fruyt, F.; McCrae, R. R.; Szirmák, Z.; and Nagy, J. 2004. The Five-Factor Personality Inventory as a Measure of the Five-Factor Model: Belgian, American, and Hungarian Comparisons with the NEO-PI-R. *Assessment* 11(3):207–215.

Eysenck, H. J. 1947. *Dimensions of Personality*. Routledge & Kegan Paul.

Eysenck, H. J. 1992. Four ways five factors are not basic. *Personality and Individual Differences* 13(6):667–673.

Golbeck, J.; Robles, C.; and Turner, K. 2011. Predicting personality with social media. In *Proceedings of Human Factors in Computing Systems (CHI'11)*, 253–262. ACM Press.

Greenwald, G. 2014. *No Place to Hide: Edward Snowden, the NSA and the Surveillance State*. Hamish Hamilton.

Gunning, R. 1968. *Technique of Clear Writing*. McGraw Hill Higher Education.

Iacobelli, F.; Gill, A. J.; Nowson, S.; and Oberlander, J. 2011. Large Scale Personality Classification of Bloggers. In *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, volume 6975 of *Lecture Notes in Computer Science*, 568–577. Springer.

Inches, G., and Crestani, F. 2012. Overview of the International Sexual Predator Identification Competition at PAN-2012. In *CLEF (Online Working Notes/Labs/Workshop)*.

Kincaid, J.; Fishburne, R.; and Chissom, R. R. B. 1975. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Technical Report Research Branch Report 8-75, Chief of Naval Technical Training: Naval Air Station Memphis.

Mairesse, F.; Walker, M. A.; Mehi, M. R.; and Moore, R. K. 2007. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research* 30:457–500.

Mairesse, F. 2013. Towards Personality-Based User Mod-

elling. Keynote talk at Workshop on Computational Personality Recognition (ICWSM-13).

Millwood-Hargrave, A. 2000. Delete expletives? Technical report, Research undertaken jointly by the Advertising Standards Authority, British Broadcasting Corporation, Broadcasting Standards Commission and the Independent Television Commission.

Mohammad, S. M., and Kiritchenko, S. 2013. Using Nuances of Emotion to Identify Personality. In *Proceedings of the ICWSM Workshop on Computational Personality Recognition*.

Norman, W. T. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* 66(6):574–583.

Oatley, G., and Crick, T. 2014. Changing Faces: Identifying Complex Behavioural Profiles. In *Proceedings of 2nd International Conference on Human Aspects of Information Security, Privacy and Trust (HAS 2014)*, volume 8533 of *Lecture Notes in Computer Science*, 282–293. Springer.

Paunonen, S. V., and Jackson, D. N. 2000. What is beyond the Big Five? Plenty! *Journal of Personality* 68(5):821–836.

Peabody, D., and Goldberg, L. 1989. Some determinants of factor structures from personality-trait descriptor. *Journal of Personality and Social Psychology* 57(3):552–567.

Pennebaker, J., and King, L. 1999. Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology* 77(6):1296–1312.

Pennebaker, J.; Francis, M.; and Booth, R. 2001. Linguistic Inquiry and Word Count. Erlbaum Publishers.

Perea, M.; Dunabeitia, J.; and Carreiras, M. 2008. R34D1NG W0RD5 W1TH NUMB3R5. *Journal of Experimental Psychology: Human Perception and Performance* 34(1):237–241.

Richter, M. M. 2003. *Readings in Case-Based Reasoning*. Morgan Kaufmann. chapter Knowledge containers.

Ritter, A.; Clark, S.; Mausam; and Etzioni, O. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*, 1524–1534.

Schiaffino, S., and Amandi, A. 2009. Intelligent User Profiling. In *Artificial Intelligence: An International Perspective*, volume 5640 of *Lecture Notes in Computer Science*, 193–216. Springer.

Smith, C.; Attwood, F.; and Barker, M. 2013. pornresearch.org Prelininary Findings. Available from: `http://www.pornresearch.org/Firstsummaryforwebsite.pdf`.

Sumner, C.; Byers, A.; Boochever, R.; and Park, G. J. 2012. Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. In *Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA 2012)*. IEEE Press.

Sweeney, K., and Whissell, C. 1984. A dictionary of affect in language: I, establishment and preliminary validation. *Perceptual and Motor Skills* 59(3):695–698.

Szirmák, Z., and De Raad, B. 1994. Taxonomy and structure of Hungarian personality traits. *European Journal of Personality* 8(2):95–117.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29(1):24–54.

The Guardian; Leigh, D.; and Harding, L. 2013. *WikiLeaks: Inside Julian Assange's War on Secrecy*. Guardian Faber Publishing.

2011. *2nd Unnatural Language Processing Contest*. part of the 17th Annual Meeting of the Association for Natural Language Processing (NLP2011): `http://www.anlp.jp/nlp2011/`.

Vazire, S., and Gosling, S. D. 2004. e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology* 87(1):123–132.

Wilson, M. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2.00. *Behavior Research Methods, Instruments & Computers* 20(1):6–10.

Woodworth, M.; Hancock, J.; Porter, S.; Hare, R.; Logan, M.; OToole, M. E.; and Smith, S. 2012. The Language of Psychopaths: New Findings and Implications for Law Enforcement. *FBI Law Enforcement Bulletin*.