Exploiting Social Media to Address Fundamental Human Rights Issues

Massimo Poesio ♠ **Ayman Alhelbawy**♣,♠ **Chris Fox** ♠ **Udo Kruschwitz**♠

♣ Minority Rights Group, London, UK

♠ University of Essex, Colchester, UK

Abstract

This invited talk provided an overview of some of our work in relation to extracting meaningful knowledge from social media feeds to help in addressing human rights issues highlighting the potential that the rise of 'big data' offers in this respect looking at both sides of the coin regarding big data and human rights: how big data can help human rights work, but also the potential dangers that can originate from the ability to analyse massive amounts of data very quickly. The primary focus of our work is on applying natural language processing methods to turn large-scale unstructured and partially structured data streams into actionable knowledge.

Keywords: Human Rights, Social Media, NLP, Arabic NLP

1. Overview

Vast amounts of social media data are being generated every second. This represents a paradigm shift in publishing from largely carefully edited data to user-generated content which, as a result, has rapidly changed the way people exchange and consume information as well as how they communicate. Managing such data streams comes with many challenges as has been discussed extensively in the research literature. Nevertheless, it also offers new opportunities. One such opportunity is the potential to more easily detect and document human rights violations. In fact, these developments have already resulted in changes to how human rights organisations work. The 'investigator on the ground' will not be completely replaced but there are many new modes of identifying evidence of human rights violations. Social media such as Facebook, YouTube and Twitter are ideal platforms to push content to the world. Obviously, there is a big challenge in validating any such postings.

Progress in natural language processing (NLP) means that off-the-shelf tools can now be used to quickly assemble a processing pipeline that takes social media data and turns it into structured knowledge. We are primarily interested in this type of processing pipeline but that needs to be seen as part of a bigger picture. Two research projects we are involved in illustrate the point.

2. Human Rights, Big Data and Technology

The *Human Rights, Big Data and Technology* (HRBDT) project[1] is an interdisciplinary research project funded by ESRC[2] based mainly at the Human Rights Centre[3] of the University of Essex with partners that include the World Health Organisation[4], the Harvard FXB Center for Health and Human Rights[5] and the Geneva Academy for International Humanitarian Law and Human Rights[6]. A core activity of one of the four workstreams is to explore and apply the potential of natural language processing to the automated analysis of 'big data' and social media and develops new approaches to humanitarian and human rights work.

3. Knowledge Transfer Partnership

The second part of our keynote talk focussed on a practical application of NLP techniques to support human rights work in a collaboration between the University of Essex and Minority Rights Group International (MRG)[7]. This project is funded by InnovateUK[8] through a Knowledge Transfer Partnership (KTP) project. The aim of this project is to provide support to civilian-led reporting of human rights violations, in the context of MRG's involvement in the Ceasefire Centre, and in particular in the Ceasefire Iraq project[9]. This project complements the objectives of the more general HRBDT project, exploring the contributions of big data – and in particular, social media – to the identification of human rights violations.

Specific objectives of the collaboration with MRG are, first of all, to develop a portal that will make it possible to collate reports of human rights violations sent by civilians using a variety of formats, from SMS to emails to social media. The portal[10], currently undergoing beta-testing and soon to go live, will allow personnel by MRG and associated organizations to view and analyse reports of human right violations sent by civilians.

Second, the project aims to develop tools to filter and analyse this type of information. The analysis techniques developed so far, and at the moment tested with tweets, include methods for detecting human rights violations reports using machine learning-based text categorization to classify text (e.g., tweets) according to a classification scheme which, in our case, includes categories such as human right violation reports (for tweets such as *"The army of Assad in Damascus committed a terrible massacre claiming the lives of dozens of children in their school"*), reporting of general violence (as in *"Four people injured as a result of a brawl in Darb Alarbaeen"*), or reporting of an accident (e.g., *"At*

[1] http://www.hrbdt.ac.uk

[2] http://www.esrc.ac.uk

[3] http://www.essex.ac.uk/hrc/

[4] http://www.who.int/

[5] https://fxb.harvard.edu

[6] http://www.geneva-academy.ch

[7] http://minorityrights.org

[8] https://www.gov.uk/government/organisations/innovate-uk

[9] http://minorityrights.org/what-we-do/ceasefire-project/

[10] http://iraq.ceasefire.org/

least 24 dead in the sinking of boat for illegal immigrants off the coast of Istanbul"). As part of the project, a dataset of over 15,000 Arabic tweets was collected and annotated according to these categories, and used to train a classifier to recognize such categories in text (Alhelbawy et al., 2016). The objective is to apply classifiers of this type to filter the data collected through the portal and/or to gather additional evidence not directly sent to the portal.

4. Conclusions

The emergence of 'big data' in the form of social media is affecting all parts of life. This development offers a lot of new challenges but also opportunities such as the application of natural language processing techniques to detect and document human rights violations. NLP tools have matured to a level that they can easily be applied, are scalable and robust. This stream of work offers the additional benefit that it applies state-of-the-art technology to practical applications that will have a measurable impact on the quality of life of many people.

Acknowledgements

The *Human Rights, Big Data and Technology* project is funded by Economic and Social Research Council grant ES/M010236/1. We also acknowledge support from InnovateUK through a Knowledge Transfer Partnership (KTP) project between MRG and the University of Essex, partnership number 9488.

5. References

Alhelbawy, A., Poesio, M., and Kruschwitz, U. (2016). Towards a corpus of violence acts in arabic social media. In *Proceedings of LREC*.