Computational Stylometry and Machine Learning for Gender and Age Detection in Cyberbullying Texts

Antonio Pascucci
UNIOR NLP Research Group
L'Orientale University of Naples
Naples, Italy
apascucci@unior.it

Vincenzo Masucci
Expert System Corp.
Naples, Italy
vmasucci@expertsystem.com

Johanna Monti
UNIOR NLP Research Group
L'Orientale University of Naples
Naples, Italy
jmonti@unior.it

Abstract—The aim of this paper is to show the importance of Computational Stylometry (CS) and Machine Learning (ML) support in author's gender and age detection in cyberbullying texts. We developed a cyberbullying detection platform and we show the results of performances in terms of Precision, Recall and F-Measure for gender and age detection in cyberbullying texts we collected.

Index Terms—Computational Stylometry, Gender Detection, Age Detection, Machine Learning, Cyberbullying Detection

I. INTRODUCTION

In this paper, we show the results of a research carried out in November 2018, during the 32nd edition of *Futuro Remoto*, the oldest European event of scientific dissemination, held in Naples by *Città della Scienza* since 1987. This research has been developed in the framework of an Innovative Industrial PhD project in CS by the "L'Orientale" University in co-operation with Expert System Corp. This research has two objectives: i) to increase the number of cyberbullying texts in our corpus for future work and ii) to demonstrate the efficiency of CS and ML in text meaning understanding.

To this aim, a cyberbullying detection web platform was developed which we tested during *Futuro Remoto*. During this event, we asked users to write texts with possible cyberbullying content, and we used the platform to detect gender and age of the writer.

This paper is organized as follows. Section II presents related work while Section III describes CS and ML. The phenomenon of cyberbullying and all the cyberbullying categories are described in Section IV. In Section V we describe our cyberbullying detection platform and the data we collected during *Futuro Remoto*. Results and conclusions are in Section VI and Future Work is discussed in Section VII.

II. RELATED WORK

As already shown in other researches, text analysis and Authorship Attribution (AA) techniques with ML algorithms support are a valid tool to oppose cyberbullying phenomena. For example, ACTS (Frommholz, et al., 2016) is a framework for automatic detection of cyberstalking texts that has a specific module based on ML algorithms trained to detect and classify cyberstalking messages.

(Yin et al., 2009) detect harassment thanks to a supervised ML approach, that takes into account the content of the harassment texts only and leaves out author's characteristics. In their study N-grams, TF-IDF score weighting and foul words frequency are used as a baseline. The results show an improvement of the baseline.

Important research has been carried out by (Bogdanova, Rosso and Solorio, 2014) and shows that understanding the behaviour of paedophiles could help to detect and prevent children sexual abuse in social media. The authors highlight that paedophiles try to be nice with a victim and make compliments, at least in the beginning of a conversation but afterwards they tend to be emotionally unstable and prone to loose temper easily and use negative words. The research is based on the following features: percentage of positive (*cute*, *pretty*) and negative (*dangerous*, *annoying*) words, percentage of sadness (*bored*, *sad*) and anger(*furious*) markers, percentage of fear (*scared*, *panic*) markers and percentage of relationship words (i.e. *boyfriend*, *date*) among others. The SVM classification based on combinations of highlevel features achieves 97% accuracy in identifying conversations with cyberpedophiles from cybersex chat logs.

(Rangel and Rosso, 2016) hypothesize that the way people write about topics expressing emotions may help to identify their demographics, their age and gender. Carrying out their experiment on Spanish texts belonging to *PAN-AP-13* corpus, they achieve an accuracy of 64% for gender detection and an accuracy of 66% for age detection. Their research is grounded in an innovative approach, where both style-based (frequency of punctuation marks, capital letters, quotations, POS tags and dictionary-based words) and emotion-labelled graphs features (that not only capture the syntactic structure but also its location in the text) are combined. This research highlights that young people tend to write more about physics, linguistics, literature, law, medicine and chemistry, maybe due to the fact that this is the stage of life when young people mostly speak about their homework. Females seem to write