2018-00636 - [CORDIS2018-WIMMICS] Cyberbullism detection in social media

Level of qualifications required: Graduate degree or equivalent
Other valued qualifications: Master degree in Data Science, Computer Science or Computational Linguistics is required.
Function: PhD Position

About the research centre or Inria department

The Inria Sophia Antipolis - Méditerranée center counts 37 research teams and 9 support departments. The center's staff (about 600 people including 400 Inria employees) is composed of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrators. 1/3 of the staff are civil servants, the others are contractual. The majority of the research teams at the center are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Six teams are based in Montpellier and a team is hosted by the computer science department of the University of Bologna in Italy. The Center is a member of the University and Institution Community (ComUE) "Université Côte d'Azur (UCA)".

Context

Hosting teams:

WIMMICS (http://wimmics.inria.fr/) is a research team of Université Côte d'Azur (UCA). The research fields of this team are graph-oriented knowledge representation, reasoning and operationalization to model and support actors, actions and interactions in web-based epistemic communities.

LJAD (http://math.unice.fr) is a laboratory of Université Côte d'Azur (UCA). The research fields of the Statistic team of LJAD are focused on network modeling, high-dimensional data clustering, functional data analysis and text analysis.

Supervisors:

- Principal advisor: Pr. Charles Bouveyron
- Co-advisors: Dr. Elena Cabrio & Dr. Serena Villata

Assignment

The rapid growth of social networking is increasing the progression of cyberbullying activities. Most of the individuals involved in these activities belong to the younger generations, especially teenagers, who in the worst scenario are at more risk of suicidal attempts. There is an urgent need to study cyberbullying in terms of its detection, prevention and mitigation. Traditional bullying is any activity by a person or a group aimed at a target group or individual involving repeated emotional, physical or verbal abuse. Bullying as a form of social problem has occurred in various forms over the years with the Web and communication technologies being used to support deliberate, repeated and hostile behaviour by an individual or group, in order to harm others (www.cyberbullying.org). Cyberbullying is defined as "an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time, against a victim who cannot easily defend him or herself" [Smith et al. 2008].

To reduce and stop cyberbullying, one effective solution is to automatically detect bullying content based on appropriate machine learning and natural language processing techniques. However, many existing approaches in the literature are just normal text classification models without considering bullying characteristics.

References:


---

**General Information**

- Theme/Domain: Data and Knowledge Representation and Processing Information system (BAP E)
- Town/city: Sophia Antipolis
- Inria Center: CRI Sophia Antipolis - Méditerranée
- Starting date: 2018-10-01
- Duration of contract: 3 years
- Deadline to apply: 2018-05-06

**Contacts**

- Inria Team: WIMMICS
- Recruiter: Cabrio Elena / elena.cabrio@inria.fr

**Conditions for application**

**Defence Security:**
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy:**
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

**Warning:** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
Main activities

In this context, the goal of this Ph.D. is to propose an effective approach to detect cyberbullying messages from social media, addressing the following two main challenges:

1. From the Natural Language Processing (NLP) perspective, the goal is to define algorithms for automatically identifying hate speech, and in particular, cyberbullying instances in short text messages. Given the high degree of dynamicity of such a kind of harmful and abusive words, adaptive strategies are required to find such words from the streaming of text. To fully capture the semantics of cyberbullying messages, the selected candidate will employ argument mining methods so that the identification of harmful and abusive words will be coupled with the identification of more subtle messages where harmful arguments are addressed against the victim of the cyberbullying attack. The identified arguments will also be classified with respect to the topic of the abuse (e.g., sex, racism). From the methodological point of view, standard machine learning algorithms like Support Vector Machines (SVM) and Logistic Regression (LR) will be used. We also plan to use deep learning methods as Recurrent Neural Networks (e.g., Long short-term memory - LSTM).

2. From the network analysis perspective, the goal is to seek structures in the network data, to identify the most active predators and victims. A possible start would be to constrain statistical models [Bouveyron et al. 2014; Nowicki et al. 2001] to discover harassment activities in social networks. To account the specificity of harassment in both terms of connectivity and language, a prior structure for the interaction between stalkers and their victims can be encoded in the model (to account for the fact that we expect that stalkers focus their communications against a few people with specific topics as hate speech sexual-related topics). This would be possible by adding constraints on the latent group structures, especially on the group proportions and the connection probabilities.

This PhD thesis proposal is particularly relevant these days given the raising importance the issue of cyberbullying is gaining. To the best of our knowledge, this is the first proposal to support the identification of cyberbullying instances in social networks using Artificial Intelligence methods. More precisely, the innovative character of the project lies in combining two different methodologies, i.e., natural language processing and network analysis, allowing to consider both the content of the messages and the links on the social network at the same time.

References:

Skills

Skills and profile:

- Master degree in Data Science, Computer Science or Computational Linguistics is required.
- Programming skills (R, Python, C++ appreciated).
- Knowledge of Natural Language Processing and Machine Learning is preferred.
- Fluent English required, both oral and written. French is appreciated but not mandatory.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)