MOVES

Monitoring Virtual Crowds in Smart Cities

Mobs, riots, and crowds greatly impact societies in numerous ways. On a mid-term basis, the collective behavior of crowds can be an agent of social change and an affirmation of existing social mores and structures. On a short-term basis, they may have dramatic consequences such as killings, slaughters, or material damages, which badly reflect how fractured our societies can be. Daily reports of demonstrations in various parts of the world emerge. Between 2012 and 2015, hundreds of thousands of people protested in the streets of Portugal against troika's public policies, sometimes resulting in violent clashes. In France, one recent anti-police brutality protest turned violent, after the death of a young environmental activist. Seven police officers were injured after around 2,000 people gathered to protest. In India, one of the recent news that attracted attention is the Gujarat violence, which happened due to the caste-based protest led by Patel Community. This caused the death of 9 people, and 18 were injured. Within this project, we propose to develop a multilingual surveillance system capable of detecting emerging crowds by identifying rising events that foster high focus, high energy and high emotion on social media. Our fundamental hypothesis is that virtual crowds evidence similar characteristics to real crowds, which may allow their modelization in terms of complex computer systems by relying on advanced natural language processing and machine learning techniques. The current project lays at the intersection of important scientific research topics, namely urban informatics, natural language processing for social media, predictive analytics over big social data and image semtiment analysis.

HOST ENTETIES

UBI
www.ubi.pt
University of Beira Interior
Covilhã, Portugal

DI
www.di.ubi.pt
Department of Computer Science
UBI

HULTIG
hultig.di.ubi.pt
Human Language Technology Information Group
UBI

SOCIALAB
socia-lab.di.ubi.pt
Soft Computing and Image Analysis Laboratory
UBI

TEXT NORMALIZATION AND ENRICHMENT

This task is commonly called named entity recognition and aims to automatically extract mentions of rigid designators from text belonging to named-entity types.
In this project, we propose to improve the robustness of named entity recognition systems for Twitter starting by following a line like the recent work of (Saha et al., 2015a) with biomedical texts. By doing so, we expect that robust classifiers can be built and used reliably in a real-word environment such as Twitter. To bridge the gap between unstructured text and structured machine readable knowledge bases, entity linking is performed and consists in mapping each entity mention in a tweet to a unique entity. A great deal of studies has been tackling entity linking (Ferragina and Scaiella, 2013) and more recently entity linking for social media texts (Liu et al., 2013).

EVENT DETECTION AND TRACKING

Social media have emerged as powerful means of communication for people looking to share and exchange information. (Vikre and Wold, 2015) show that the use of locality-sensitive hashing combined with named entity recognition achieves better performance for detecting news than using the topic modeling approach. Moreover, topics are represented as sets of words and thus may include more general topics than specific ones.
Therefore, we propose a new strategy based on the recent findings of (Moreno et al., 2014) which proposed the Dual C-means clustering algorithm allowing to mix document-pivot and feature-pivot techniques into a unique model. Moreover, the clustering process can be driven by bursty keywords or named entities, instead of relying on all possible terms present in the time window.

RESEARCHERS TEAM

R&D Leaders

Sebastião Pais, PhD
Statistical Natural Language Processing
Lexical Semantics

João Paulo Cordeiro, PhD
Natural Language Processing
Automatic Text Summarization

Luís Alexandre, PhD
Pattern Recognition
Image Processing

EXTREMISM AND COLLECTIVE RADICALIZATION UNDERSTANDING

Indeed, ultimately, a crowd is characterized by its dominant emotion, its level of interaction and shared focus. (Krumm, 2015) showed that specific radicalized language is used within acting and protest crowds. As such, we hypothesize that radicalized language mainly expresses negative emotions (such as anger, fear, or anxiety) with high intensity, following the classification of Plutchik's wheel of emotion (Plutchik, 1980).
In this project we propose to learn weak classifiers based on dictionaries of emotional words (Qadir and Riloff, 2014) to retrieve many roughly classified emotional tweets, in a similar way as (Tang et al., 2014). For that purpose, weak classifiers for intensity detection will be built based on recent work of (Sharma et al., 2015) on language intensity.
Finally, we will study the introduction of demographically-driven word embeddings.

IMAGE SENTIMENT ANALYSIS

Extracting sentiment information from images is a hard task given the fact that the same image can be interpreted by different people as conveying a different sentiment. Nonetheless, several approaches to Image Sentiment Analysis (ISA) have been proposed and there are already commercial services that implement ISA.
First, we intend to use deep learning approaches (Goodfellow et al., 2016), to model sentiment in images from pre-labeled databases, such as (Borth et al. 2013). Second, we will use methods for auto-labeling of images, such as ((Vinyals et al. 2015), (Karpathy et al., 2017)), and infer the sentiment from the produced labels. Thirdly, we will take advantage of the sentiments in the text that accompanies the images to automatically label a large dataset of tweets and then train the deep learning models using this dataset.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

COMPETE 2020

PORTUGAL 2020

E-mail
moves@di.ubi.pt

Website
www.moves.di.ubi.pt

Address
University of Beira Interior
Department of Computer Science
Rua Marquês d'Ávila e Bolama
6201-001 Covilhã
Portugal
+351 275 242 081