

Workshop on Machine Learning for Trend and Weak Signal Detection in Social Networks and Social Media



<https://www.irit.fr/twsdetection>

Introduction	3
1 Objectives	4
2 Organisation Committee :	5
3 Program	6
Session1: Agressive text detection	6
Session 2: Automatic detection	6
Session 3: Ethic and privacy issue	7
Session 5: Fake news detection	7
Session 1 : Aggressive text detection	8
1 The Role of Storylines in Hate Speech Detection	9
2 Hatemeter Project: Analysis of hate speech on twitter at the crossroads of computer science, humanities and social sciences	10
3 Aggression Identification in Posts - two machine learning approaches.	11
Session 2 : Automatic detection	12
1 Prediction and Visual Intelligence Platform for Detection of Irregularities and Abnormal Behaviour	13
2 Deep learning with weakly-annotated data: a sound event detection use case	14
3 Topical Community Detection: an Embedding User and Content Similarity Method	15
4 What's in the News? Identification of Trending Topics in Alternative and Mainstream Lithuanian Media	16
Session 3 : Ethic and privacy issue	17
1 Ethical and legal challenges of machine learning for trend and weak signal detection in social networks: an overview	18
2 Media Ethics and Education of the Media within the framework of social media and social networks	19
3 Intelligence analysis and semantic interoperability for identity resolution	20

Session 4 : Social Media : Radicalisation and Misinformation.....	21
Targeting Societal Challenges by analysing social media data.	22
Session 5: Fake news detection.....	23
1 Detecting fake news in social media content	24
2 Information Nutritional Label to Predict Information Check-Worthiness 25	
Sponsors	26

Introduction

1 Objectives

The objective of this workshop is to provide an opportunity for researchers and practitioners from different disciplines to meet, exchange and discuss issues, research results and practical applications regarding trend and weak signal detection in social networks. This workshop is also a good chance for participants to develop their network and to cooperate with others.

This workshop had taken place on February 27-28, 2020 , Toulouse, France at INSPE, 56 avenue de l'URSS, 31400 Toulouse.

More information at <https://www.irit.fr/twsdetection/>.

This workshop invited submissions covering the following topics, but not limited to:

- Trend detection, analysis and tracking
- Topic identification and event detection
- Weak signal detection
- Information / opinion / knowledge spread and modelling
- Misinformation and misbehavior analysis and detection
- Information quality in social network
- Community detection, expertise and authority discovery
- Social influence, recommendation and media
- Behavior analysis in social networks
- Sentiment analysis
- Crime detection and investigation
- Network visualization and modeling
- Data mining and machine learning
- Real-world case studies
- Ongoing projects based on social media and/or social networks.

2 Organisation Committee :

Thi Bich Ngoc Hoang, Institut de Recherche en Informatique de Toulouse (IRIT), UMR5505 CNRS, France, chair

Adrienne Gaultier, IRIT, UMR5505, Toulouse

Pascal Marchand, Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales (LERASS), l'Université Paul Sabatier, France

Béatrice Milard, Interdisciplinaire Solidarités, Sociétés, Territoires (LISST) Toulouse, France

Josiane Mothe, IRIT, UMR5505 CNRS, INSPE-UT2, Université de Toulouse, France

Md Zia Ullah, IRIT, UMR5505 CNRS, France

3 Program

Session1: Agressive text detection

The Role of Storylines in Hate Speech Detection,

Kurt Englmeier,

Hochschule Schmalkalden Fakultät Informatik Blechhammer 4-9 98574
Schmalkalden

- **Hatometer Project: Analysis of hate speech on twitter at the crossroads of computer science, humanities and social sciences,**
Mario Laurent,
Projet Hatometer Université Toulouse 1 Capitole / MSHS-T
- **Aggression Identification in Posts - two machine learning approaches,**
Faneva Ramiandrisoa,
IRIT & Université de Toulouse, France

Session 2: Automatic detection

- **Prediction and Visual Intelligence Platform for Detection of Irregularities and Abnormal Behaviour,**
Konstantinos Demestichas, Theodoros Alexakis, Nikolaos Peppes, Konstantina Re-moundou, Ioannis Loumiotis, Wilmuth Müller, Konstantinos Avgerinakis
Institute of Communication and Computer Systems, 9, Iroon. Polytechniou Str., 15773 Zografou Athens, Greece
- **Deep learning with weakly-annotated data: a sound event detection use case,**
Thomas Pellegrini,
Assistant Professor in Computer Science, IRIT, Toulouse, France
- **Topical Community Detection: an Embedding User and Content Similarity Method,**
Thi Bich Ngoc Hoang,
IRIT, UMR5505 CNRS, France University of Economics, the University of Danang, Vietnam

- **What's in the News? Identification of Trending Topics in Alternative and Mainstream Lithuanian Media**
Justina Mandravickaitė, Monika Briedienė¹, Jonas Uus and Tomas Krilavičius
 Baltic Institute of Advanced Technology, Pilies str. 16, Vilnius 01124, Lithuania
 Vytautas Magnus University, K. Donelaičio str. 58, Kaunas 44248, Lithuania

Session 3: Ethic and privacy issue

- **Ethical and legal challenges of machine learning for trend and weak signal detection in social networks: an overview,**
Zuzanna Warso,
 Trilateral Research, London
- **Media Ethics and Education of the Media within the framework of social media and social networks,**
Sabahudin Hadžialić and Thi Phuong Vi,
 UNINETTUNO University, Rome, Italy
 Thai Nguyen University, Vietnam
- **Privacy preserving intelligence analysis for resolving identities,**
Kostas Davarakis & Eva Blomqvist,

Session 4: Social media: Radicalisation and Misinformation

- **Targeting Societal Challenges by analysing social media data,**
Miriam Fernandez
 Senior Research Fellow at the Knowledge Media Institute

Session 5: Fake news detection

- **Detecting fake news in social media content,**
Charles Huot and Sonia Collada Pérez,
- **Information Nutritional Label to Predict Information Check-Worthiness,**
Md Zia Ullah
 IRIT, UMR5505 CNRS, 118 Route de Narbonne, 31062 Toulouse CEDEX 9,
 France

Session 1 : Aggressive text detection

1 The Role of Storylines in Hate Speech Detection

Kurt Englmeier

Schmalkalden University of Applied Science, Blechhammer, 98574 Schmalkalden,
Germany
kurtenglmeier@acm.org

Abstract. The paper demonstrates the use of named-entity recognition for automatic hate speech detection. Our approach explicitly also addresses the design of models to map storylines and social anchors. They provide valuable background information for the analysis and correct classification of the brief statements used in social media. Named-entity recognition may in any case help to tackle the specifics of the language style often used in hate tweets, a style that differs from regular language in deliberate and unintentional misspellings, strange abbreviations and inter-punctuations, and the use of symbols. We implemented a prototype for our approach that automatically analyzes tweets along storylines. It operates on a series of bags of words containing names of persons, locations, characteristic words for insults, threats, and phenomena reflected in social anchors as the social background of tweets. We demonstrate our approach using a collection of German tweets that refer to the vitally discussed topic “refugees” in Germany.

Keywords: Hate Speech Detection, Named-Entity Recognition, Social Anchor, Storyline.

2 **Hatometer Project: Analysis of hate speech on twitter at the crossroads of computer science, humanities and social sciences**

Mario Laurent¹

¹ Université Capitole, Toulouse, France

Abstract. The European project "Hatometer" aims to better understand, acknowledge and prevent online hate speech on social media. The project was established through a partnership between three universities, a research institute, and three NGOs committed to the defense of human rights, based in Italy, France and the United Kingdom. To begin with, our multidisciplinary team (criminology, social sciences, computer sciences and law) focused on hatred directed against Muslims on Twitter. Supported by on the field expertise of NGOs, we developed an ICT tool (i.e., Hatometer Platform) that automatically monitors and analyses social media data on the phenomenon. The Platform allows its users to: identify potential hate messages in real-time; visualize data collected through hashtag, or user, co-occurrence graphs; observe the evolution of targeted speech over time; and propose counter-narratives examples. Each partner has identified advantages and limitations to the platform. From a research point of view, it has proven to be a good tool to run a quantitative exploration, identify groups and discourses in order to carry out a more detailed analysis of their exchanges. For NGOs, while the platform alone is not sufficient to nuance stereotypes, to curb the spread of hate speech or fake news, it seems useful to stimulate and speed up the training of operators in charge of producing counter-narratives on social media.

Keywords: Online hate Speech, Anti-Muslims hatred, Twitter monitoring, NLP applied on Social Sciences.

3 Aggression Identification in Posts - two machine learning approaches.

Faneva Ramiandrisoa
IRIT & Université de Toulouse, France
faneva.ramiandrisoa@irit.fr

Abstract. Social media have changed the way people communicate in many ways. One of the aspects is cyber-aggression and interpersonal aggression that can be catalyzed by perceived anonymity. Automatically monitoring user-generated content in order to help moderating this content is thus a hot topic. In this paper, we present and evaluate two supervised machine learning models to identify aggressive content and the level of aggressiveness. The first model uses random forest and linear regression while the second model uses deep learning techniques.

Keywords: Social media; Social media analysis; Cyber-aggression; TRAC Trolling, Aggression and Cyberbullying; Machine learning based model

Session 2 : Automatic detection

1 **Prediction and Visual Intelligence Platform for Detection of Irregularities and Abnormal Behaviour**

Konstantinos Demestichas¹, Theodoros Alexakis, Nikolaos Peppes, Konstantina Re-moundou, Ioannis Loumiotis, Wilmuth Müller, Konstantinos Avgerinakis

¹Institute of Communication and Computer Systems, 9, Iroon.Polytechniou Str.,
15773 Zografou Athens, Greece
cdemest@cn.ntua.gr

Abstract. Nowadays, (cyber)criminals demonstrate an ever more increasing re-solve to exploit new technology so to achieve their unlawful purposes. There-fore, Law Enforcement Agencies (LEAs) should accommodate an approach that surpass the existing limits in policing practices. In this light, the authors intro-duce an innovative platform that provides near real-time advanced social behav-ior analytics using irregularities detection based on historical patterns.

Keywords: Big Data, abnormal behaviour detection, crime detection.

2 Deep learning with weakly-annotated data: a sound event detection use case

Thomas Pellegrini
IRIT & Université de Toulouse, France
thomas.pellegrini@irit.f

Abstract. Weakly-annotated data correspond to data manually annotated with "weak" labels. Weak labels refer to global tags, at document level, with no information about the precise location (in time or space) of the events of interest. Deep neural networks can be trained with these data as predictors of the tags of interest. We would like to design methods to go further by trying to use these networks to also predict where the events of interest are localized within the input data. Weakly-supervised deep learning approaches will be described, with sound event detection as a use case. I will review two main research directions: i) the introduction of attention mechanisms in the network architecture, ii) the use of Multiple Instance Learning inspired objective functions. I will comment on their limitations and how these could be overcome.

Keywords. weakly-annotated data, lightly-supervised deep learning, sound event detection

3 **Topical Community Detection: an Embedding User and Content Similarity Method**

Thi Bich Ngoc Hoang^{1:2}

¹ IRIT, UMR5505 CNRS, France

² University of Economics, the University of Danang, Vietnam

Abstract. Community detection aims at partitioning a network into subgroups of densely connected nodes. While many approaches focus on community detection based on users' relationships, the latter may be not effectively enough since some communities may be topic dependent. In this paper, we propose a method that detects communities by considering users and their topics. More specifically, our approach combines cues extracted from the users' exchanges and the ones extracted from their posts. The data collection and the evaluation measures we intend to apply our method on are also presented in this paper.

Keywords. Social Media Analysis, Community Detection, Embedding User, Content Similarity, Twitter

4 What's in the News? Identification of Trending Topics in Alternative and Mainstream Lithuanian Media

Justina Mandravickaitė¹, Monika Briedienė^{1,2}, Jonas Uus¹ and Tomas Krilavičius^{1,2}

¹ Baltic Institute of Advanced Technology, Pilies str. 16, Vilnius 01124, Lithuania

² Vytautas Magnus University, K. Donelaičio str. 58, Kaunas 44248, Lithuania
justina@bpti.lt

Abstract. Tracking topics dynamics and focus in different media channels is an important tool for opinion-forming mechanisms and process analysis. Information collect, text analytics and Artificial Intelligence tools allows identification of trending topics in different media sources, while exploratory visual analytics tools provide means to identify prevalence of topics in different sources, and their dynamics. In this paper we discuss an ongoing research and demonstrate applicability of such approach to main Lithuanian news portal (delfi.lt) and one of the main alternative media channels – sarmatas.lt.

Keywords: Topic modelling, Framing, NLP, Lithuanian language, Artificial Intelligence.

Session 3 : Ethic and privacy issue

1 Ethical and legal challenges of machine learning for trend and weak signal detection in social networks: an overview

Zuzanna Warso,

Trilateral Research

Abstract. This presentation is meant to provide an overview of some of the ethical and legal challenges arising from the use of machine learning for trend and weak signal detection in social networks. While it does not exhaust the topic, the presentation's aim is to raise awareness among researchers carrying out this type of activities in order to ensure compliance with ethical and legal standards related to personal data protection applicable in the European Union. After providing an overview of basic concepts, it discusses the following issues related to personal data protection: the use of data from social networks (social media), the processing of special categories of data and the question of "explainability".

Keywords: ethics, privacy, big data ethics, GDPR, personal data protection, social media data, open data, explainability

2 **Media Ethics and Education of the Media within the framework of social media and social networks**

Sabahudin Hadžialić¹ and Thi Phuong Vi²

¹Assoc. Prof. Dr. and Dr., UNINETTUNO University, Rome, Italy,

²Ph. D Student, Thai Nguyen University, Vietnam

Abstract With outstanding advantages, social media has been changing the habit of searching, sharing and using of the public media. The problem is that, in an "open" society, social media is often associated with informal communication activities, because it allows users to easily join a certain group on social networks to chat and make your own point of view on an issue they are concerned about. Of course, to be able to use and not to be abused by social media, we need a media literate, open-minded people who are critically involved in this kind of the creation and exchange of media content. From the technical art of media perspective, social media is operated based on online services, news can be shared and spread quickly and interactively among people. join. The top issue is to attach importance to media ethics in social media.

Keyword: education, media, ethics, social media, media literacy

3 Intelligence analysis and semantic interoperability for identity resolution

Eva Blomqvist¹, Dr. Costas Davarakis²

¹Department of Computer and Information Science, Linköping University,

²Senior Consultant - Research Coordinator, SPIRIT Technical Coordinator

Abstract. SPIRIT tries to define habitual patterns as social graph related entity patterns that tend to re-appear as the perpetrators attempt to perform identity hiding movements. The goal of the SPIRIT tools identity resolution service is to help break the perpetrator's tendency to resort to similar social graph related entity patterns, by bringing a certain external awareness of these patterns. Moreover, branching away from the natural idea of perpetrators habitual patterns where there is a hazy distinction between ethics preserving identity hiding behaviours and outlaw behaviour, the process of the SPIRIT identity resolution service is designed to learn the patterns and feed it (also visually) to the social graph related entity patterns to enable identity resolution investigation jobs.

The SPIRIT tools are developed in an implementation loop. The statement of overall general requirements from the End User partners (LEAs) clearly articulates the challenge that has to be solved by the SPIRIT architecture; To enhance the effectiveness and efficiency of LEAs to tackle crime, disorder and terrorism through the delivery of a toolkit to utilise the combined capabilities of one or more open source search engines to provide a wider and richer return for specified entities, presenting the results in a manner that supports further refined search or a final product that has been subject to (semi) automated categorisation, visualisation and verification. The toolkit must also comply with existing data protection and security protocols and be capable of receiving automated products from existing in-house Police systems to (semi) automatically produce a list of potential identities who may belong to a single person.

In this presentation we will discuss first the main task of entity resolution, and the SPIRIT prototypes that will realise the solution. Then we will discuss the ethical and privacy issues targeted by the project, and finally discuss also a specific technical challenge, i.e. semantic interoperability when integrating SPIRIT data.

Keywords. Identity resolution, Social graph, Semantic interoperability, Privacy, Ethics

**Session 4 : Social Media : Radicalisation and
Misinformation**

Targeting Societal Challenges by analysing social media data.

Miriam Fernandez

Senior Research Fellow at the Knowledge Media Institute

Bio. Dr Miriam Fernandez is a Senior Research Fellow at the Knowledge Media Institute (KMi), Open University, and a member of KMi's Social Data Science Group. Before joining KMi, she was research associate at Universidad Autonoma de Madrid, Spain and software engineer (internship) at Google Zurich, Switzerland. Her research is at the intersection of the Web Science (WS) and Semantic Web communities (SW), where she has contributed with more than 100 peer-reviewed articles in various leading conferences and journals. She has extensive expertise in leading EU and national projects. She frequently participates in organising committees and editorial boards of the top SW and WS conferences, recently being program co-chair of the International Semantic Web conference in 2017, and serving as editor for the Journal of Web Semantics. Dr. Fernandez is the Athena SWAN champion for KMi, leading key initiatives to increase equality diversity in higher education.

Abstract. The Social Web has become one of the largest human information and communication systems in history, impacting the lives of billions of people around the world. This social phenomenon is transforming the world in ways that were never imagined, shaping how we communicate and share information, how we engage with others, or how we elect governments. In this talk we discuss how, by mining this unprecedented resource, we can help targeting societal challenges, with a particular focus on the problems of radicalisation and misinformation.

Session 5: Fake news detection

1 Detecting fake news in social media content

Charles Huot, Sonia Collada

Expert System

Abstract. SocialTruth is an European project designed and developed by an international consortium of 11 partners, whose slogan is “Embedding veracity for social media and web”. Creating a system that ranks news and certifies their reliability, targeted to professionals and ordinary users of social media and the web.

Expert System has conceived a system providing textual and semantic analysis; more concretely a meta-verification system on story classification and ranking. The global aim of the approach we have worked on is, based on a golden corpus, comparing an untargeted document to our base of “qualified documents”. The hypothesis applied lies in the fact that “true news” have a pattern that our system will highlight, being able to go from a large group of articles to automatically select a relevant group to compare to a possible “fake news”.

The system defined goes through the following milestones: The first milestone is the categorization, having the untargeted document; we analyze it using the categorization process and reduce the list to compare it to. The second milestone corresponds to narrowing the search by using the clustering, which means applying an unsupervised categorization among a whole corpus. The third milestone is the similarity analysis: how similar this document is to other equivalent documents? The fourth milestone corresponds to sentiment analysis, whether there is a strong sentiment present or not in the text and the writing style. Finally, we end up by extracting the writeprint of the document, language level, tone and type of vocabulary.

The outcome of our work is the conception of a semantic analyzer that aims at providing information that will be the input of the expert meta-verification system. The expert meta-verification system will combine the verification results from the content verification services created from social, semantic and multimedia content, in order to compute a meta-score that accurately depicts the credibility of the digital content under consideration.

Keywords. Social media and social networks, Content verification, Semantic analysis, Machine learning, Expert system, Fake news

2 Information Nutritional Label to Predict Information Check-Worthiness

Md Zia Ullah

IRIT, UMR5505 CNRS

mdzia.ullah@irit.fr

118 Route de Narbonne, 31062 Toulouse CEDEX 9, France

Abstract. In this communication, we introduce the important problem of information check-worthiness. We present the method we developed to automatically answer this problem. This method makes use of an elaborated information representation that combines the “information nutritional label” features along with word-embedding features. The information check-worthy claim is then predicted by training a machine learning model based on these features. Our model outperforms the official participants’ runs of CheckThat! 2018 challenge.

Keywords: Information check-worthiness; Information nutritional label; Machine learning based model.

Sponsors



Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J



This workshop is partially supported by the European Union's Horizon 2020 H2020-SU-SEC-2018 under the Grant Agreement n°833115 (PREVISION project).