

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

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



Toulouse, February 27-28, 2020

© Nguyen Gia Hung

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

1 Objectives of the workshop

The objective of this workshop was 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 was an opportunity for participants to develop their network and to cooperate with each other.

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

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.

We called for contributions that could be either unpublished research results as well as project presentations and ongoing research. All the papers that were submitted have been peer reviewed by three reviewers. Since the main idea of the workshop was to share science, ongoing research without strong result yet also could be accepted. Out of the scope papers have been rejected.

Some papers have been accepted as long papers, some other are short papers. Finally, we invited some researchers to give a talk. For those, we publish here the abstract of their talk as well as a short bio.

2 Committees

2.1 Organisation Committee

The organisation committee was a multi-disciplinary, implying researchers in computer science, sociologists.

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

Pascal Marchand, Information and communication sciences
Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales
(LERASS), Université Paul Sabatier,
Toulouse, France

Béatrice Milard, Sociologist
Interdisciplinaire Solidarités, Sociétés, Territoires (LISST), UMR 5193 CNRS
Toulouse, France

Josiane Mothe, Computer scientist
Institut de Recherche Informatique de Toulouse (IRIT), UMR5505 CNRS,
INSPE-UT2, Université de Toulouse,
Toulouse, France

Adrienne Gaultier,
Institut de Recherche en Informatique de Toulouse (IRIT), UMR5505 CNRS,
Toulouse, France

Md Zia Ullah, Computer scientist
Institut de Recherche en Informatique de Toulouse (IRIT), UMR5505 CNRS,
Toulouse France

2.2 Program Committee

The international program committee was composed of:

- **Nadine Baptiste-Jessel**
IRIT, UMR5505 CNRS,
France
- **Adrian-Gabriel Chifu**,
University of Aix-Marseille,
France

- **Michał Choraś,**
University of Science and Technology,
Bydgoszcz, Poland
- **Fabio Crestani,**
University of Italian Switzerland,
Switzerland
- **Liana Ermakova,**
University of Bretagne Occidentale,
France
- **Tomas Krilavičius,**
Baltic Institute of Advanced Technology,
Vilnius, Lithuania
- **Marion Maisonobe,**
UMR Géographie-cités,
Paris, France
- **Véronique Moriceau,**
IRIT, UMR5505 CNRS,
France
- **Quốc-Vinh Nguyen-Tran,**
Danang University of Science and Education,
Vietnam
- **Francisco-José Pérez-Carrasco,**
Technical University of Valencia,
Spain
- **Eric SanJuan,**
University of Avignon,
France

We would like to thank them all for their work and their reviews.

3 Overview of the proceeding content

The current proceedings follow the workshop program and is organized into sessions as the workshop was. Each session corresponds to a topic that was discussed during the workshop.

The first session focuses on “**Risk and Trend Detection**”.

Social media eases rapid connection of people and quick and huge diffusion of contents. It is also known to be a mean for radical group to communicate. This first session focuses on the associated challenges. It consists in four contributions. The first contribution comes from an invited speaker, Mirian Fernandez, Senior Research Fellow at the Knowledge Media Institute, Open University. In her talk untitled “*Targeting Societal Challenges by analysing social media data*”, she presented different parts of her and colleagues’ research results and targeted mainly radicalization detection in social media. The abstract of her talk and bio are included in the proceedings. Tomas Krilavičius, Baltic Institute of Advanced Technology, and Vytautas Magnus University, Lithuania presented the second communication untitled “*What’s in the News? Identification of Trending Topics in Alternative and Mainstream Lithuanian Media*” where a deep analysis was developed. A long paper is associated to this presentation. The third contribution is from Thi Bich Ngoc Hoang from IRIT-CNRS and is untitled “*Topical Community Detection: an Embedding User and Content Similarity Method*” where she presents a model to detect communities that not only considers the interactions between actors but also the content they exchange. The final contribution is this first session has been presented by Konstantinos Demestichas, member of the Institute of Communication and Computer Systems, Greece where he presented a platform developed within the PREVISION European project; the short related paper is untitled “*Prediction and Visual Intelligence Platform for Detection of Irregularities and Abnormal Behaviour*”.

The second session focuses on “**Detection of Aggressive Content**”.

Social media allow people to hide their identity and this has led to the proliferation of abusive language and increase of aggressive content on social media. Automatically identifying specific content as for example aggressive one is thus a hot research topic.

The four communications in this session are as follows. Thomas Pellegrini, researcher at the IRIT CNRS lab, France was an invited speaker. He presented some of his research work in his presentation untitled “*Deep learning with weakly-annotated data: a sound event detection use case (and hate speech detection here and there)*”. He also presented a very accurate state of the art of deep learning methods and above all on transparency, soft et self-attention -based models, multiple instances and semi-supervised learning. The abstract of his talk and bio are included in the proceedings. The two next communications were both tackling the same challenge of hate speech or content detection. Kurt Englmeier from Hochschule Schmalkalden Fakultät Informatik Blechhammer, Germany presented the his contribution untitled “*The Role of Storylines in Hate Speech Detection*”. Faneva Ramiandrisoa, from IRIT CNRS lab, presented his PhD work on “*Aggression Identification in Posts - two machine learning approaches*”.

The two corresponding long paper are included in the proceeding. The final contribution was from Mario Laurent, MSHS-T, Université de Toulouse where he presented the “*Hatemeter Project: Analysis of hate speech on twitter at the crossroads of computer science, humanities and social sciences*”. This is a H2020 funded project where the idea is to go a step further after hate speech detection by helping the development and spreading of counter speeches. A long paper is also associated to this communication.

The third session is on **Fake news detection**.

Social media makes information spreading easier and quicker, but in many cases regardless of the information quality. Automate fact-checking has recently become a hot topic.

Two communications have been presented within this session. Md Zia Ullah from IRIT CNRS lab presented a communication untitled “*An ML Model for Predicting Information Check-Worthiness using a Variety of Features*” associated with a short paper in these proceedings. Charles Huot and Sonia Collada Pérez from Expert System presented a communication on the same topic untitled “*Detecting fake news in social media content*”

The fourth and final session focused on **Ethics and privacy issue**.

Social data analysis raises ethical and privacy issues. Two communications were presented on this topic. As an invited speaker, Zuzanna Warso, from Trilateral Research, London gave a clear and completed overview on this issue, untitled “*Ethical and legal challenges of machine learning for trend and weak signal detection in social networks: an overview*”. Included in these proceedings is an abstract related to her presentation and her bio. Finally, Kostas Davarakis & Eva Blomqvist, from Sigular Logic, Linkoping university presented the latest communication that is included in this proceeding under the form of a abstract untitled “*Privacy preserving intelligence analysis for resolving identities*”