LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

Workshop on Automated Event Extraction of Socio-political Events from News (AESPEN2020)

PROCEEDINGS

Editors:
Ali Hürriyetoğlu (Koc University)
Erdem Yörük (Koc University and University of Oxford)
Hristo Tanev (European Commission Joint Research Center)
Vanni Zavarella (European Commission Joint Research Center)

Proceedings of the LREC 2020 Workshop on Automated Event Extraction of Socio-political Events from News (AESPEN2020)

Edited by: Ali Hürriyetoğlu, Erdem Yörük, Hristo Tanev, and Vanni Zavarella

ISBN: 979-10-95546-50-4
EAN: 9791095546504

For more information:
European Language Resources Association (ELRA)
9 rue des Cordelières
75013, Paris
France
http://www.elra.info
Email: lrec@elda.org

Introduction

This year we've accepted for publication nine papers which look at event detection from different points of view.

Nine papers were regular paper submissions and one was a shared task participation report. The shared task report and seven of the regular papers were accepted on the basis of reviews, which were five per paper, performed by the program committee members.

The accepted regular papers can be grouped as i) evaluation of state-of-the-art machine learning approaches by Buyukoz et al., Olsson et al., and Piskorski, and Jacquet, ii) introduction of a new data set by Radford, iii) projects of event information collection by Osorio et al. and Papanikolaou and Papageorgiou, and iv) forecasting of political conflict by Halkia et al.

The evaluation of Buyukoz et al. and Olsson et al. show that state-of-the-art deep learning models such as BERT and ELMo (Peters et al., 2018) yield consistently higher performance than traditional ML methods such as support vector machines (SVM) on conflict and protest event data respectively. Piskorski and Jacquet have found that TF-IDFweighted character n-gram based SVM model performs better than an SVM model that facilitates pre-trained em-beddings such as GLOVE (Pennington et al., 2014), BERT, and FASTTEXT (Mikolov et al., 2018) in most of the experiments on conflict data. Radford introduces the dataset Headlines of War for cross-document coreference resolution for the news headlines.The dataset consists of positive samples from Militarized Interstate Disputes dataset and negative samples from NewYork Times. The description of this invaluable resource accompanied with a detailed discussion of its utility and caveats. Osorio et al. introduce Hadath that is a supervised protocol for event information collection from Arabic sources. The utility of Hadath was demonstrated in processing news reported between 2012 and 2012 in Afghanistan. In the scope of the other event information collection study, Papanikolaou and Papageorgiou processed two news sourcesi in Greek from Greece to create a database of protest events for the period between 1996 and 2014. Osorio et. al. and Pa-panikolaou and Papageorgiou utilized fully automatic toolsthat integrates supervised machine learning and rule based methodologies at various degrees. Finally, Halkia et al. presents a material conflict forecastingstudy that facilitates the available event databases GDELT and ICEWS. Their results demonstrate that it is possible to correctly predict social upheaval using the methodology they propose, which utilizes Long-Short Term Memory(LSTM) (Hochreiter and Schmidhuber, 1997).We have received expression of interest from 12 research teams, of which 6 teams signed the application form and received the data. Two of these teams sent their predictions on the test data. Finally, only Ors et al. submitted a paper about their work. This team reported their work as consisting of three steps. First, they use a transformer based model,which is ALBERT (Lan et al., 2020), to predict whether apair of sentences refer to the same event or not. Later, they use these predictions as the initial scores and recalculate their scores by considering the relation of sentences in a pair with respect to other sentences. As the last step, final scores between these sentences are used to construct the clusters,starting with the pairs with the highest scores.

The variety of the submitted papers show that we could bring the ML, NLP, and social and political science communities together. Although the breadth of the topics were limited, the technical depth and timeliness of the contributions show that the workshop contribute to the discipline of automatic extraction of socio-political events. The papers about processing Arabic and Greek sources are significant contributions to the understanding of how should we handle languages other than English. Finally, the shared task ESCI demonstrated the prevalence of the event coreferences, some baselines for handling them, and a state-of-the-art system that is able to tackle this task.

Organizers:

Ali Hürriyetoğlu (Koc University)
Erdem Yörük (Koc University and University of Oxford)
Hristo Tanev (European Commission – Joint Research Center)
Vanni Zavarella (European Commission – Joint Research Center)

Program Committee:

Svetla Boycheva (Institute of Information and Communication Technologies, Bulgarian Academy of Sciences)
Fırat Duruşan (Koc University)
Theresa Gessler (University of Zürich)
Christian Göbel (University of Vienna)
Burak Gürel (Koc University)
Matina Halkia (European Commission – Joint Research Center)
Sophia Hunger (European University Institute)
J. Craig Jenkins (The Ohio State University)
Liron Lavi (UCLA Y&S Nazarian Center for Israel Studies)
Jasmine Lorenzini (University of Geneva)
Bernardo Magnini (Fondazione Bruno Kessler (FBK))
Osman Mutlu (Koc University)
Nelleke Oostdijk (Radboud University)
Arzucan Özgür (Boğaziçi University)
Jakub Piskorski (Polish Academy of Sciences)
Lidia Pivovarova (University of Helsinki)
Benjamin J. Radford (UNC Charlotte)
Clionadh Raleigh (University of Sussex)
Ali Safaya (Koc University)
Parang Saraf (Virginia Tech)
Philip Schrodt (Parus Analytical Systems)
Manuela Speranza (Fondazione Bruno Kessler, Trento)
Çağrı Yoltar (Koc University)
Aline Villavicencio (The University of Sheffield)
Kalliopi Zervanou (Eindhoven University of Technology)

Invited Speakers:

Clionadh Raleigh, University of Sussex
Philip Schrodt (Parus Analytical Systems)

Table of Contents

Workshop Program

Automated Extraction of Socio-political Events from News (AESPEN): Workshop and Shared Task Report
Ali Hürriyetoğlu, Vanni Zavarella, Hristo Tanev, Erdem Yörük, Ali Safaya and Osman Mutlu

Keynote Abstract: Too soon? The limitations of AI for event data
Clionadh Raleigh

Keynote Abstract: Current Open Questions for Operational Event Data
Philip A. Schrodt

Analyzing ELMo and DistilBERT on Socio-political News Classification
Berfu Büyüköz, Ali Hürriyetoğlu and Arzucan Özgür

Text Categorization for Conflict Event Annotation
Fredrik Olsson, Magnus Sahlgren, Fehmi ben Abdesslem, Ariel Ekgren and Kristine Eck

TF-IDF Character N-grams versus Word Embedding-based Models for Fine-grained Event Classification: A Preliminary Study
Jakub Piskorski and Guillaume Jacquet

Seeing the Forest and the Trees: Detection and Cross-Document Coreference Resolution of Militarized Interstate Disputes
Benjamin Radford

Conflict Event Modelling: Research Experiment and Event Data Limitations
Matina Halkia, Stefano Ferri, Michail Papazoglou, Marie-Sophie Van Damme and Dimitrios Thomakos

Supervised Event Coding from Text Written in Arabic: Introducing Hadath
Javier Osorio, Alejandro Reyes, Alejandro Beltrán and Atal Ahmadzai

Protest Event Analysis: A Longitudinal Analysis for Greece
Konstantina Papanikolaou and Haris Papageorgiou

Event Clustering within News Articles
Faik Kerem Örs, Süveyda Yeniterzi and Reyyan Yeniterzi