Call for Shared Task Participation at CASE @ ACL-IJCNLP: Socio-political and Crisis Events Detection

Event information detection consists of multiple subsequent steps that could drastically affect the quality of the resulted event database. Thus, we believe one must consider a complete scenario that consists of document and sentence classification as relevant or not, event coreference resolution, event information extraction, and event classification in relation to an event taxonomy, and test the results on a list of events created manually to determine performance of the state-of-the-art on this task.

With this objective in mind, we organize a shared task on socio-political and crisis event detection at the workshop CASE @ ACL-IJCNLP 2021 (https://emw.ku.edu.tr/case-2021/). Although the subtasks form a coherent flow, task participants can focus on one or more of them. Therefore, participants can choose the tasks or subtask(s) they would like to participate in. Participants will have access to all of the data for all tasks and subtasks. Any combination of these resources to achieve high performance for any of the tasks is allowed. The tasks and subtasks are:

Task 1. Multilingual protest news detection

- Subtask 1: Document classification
  - Does a news article contain information about a past or ongoing event?
- Subtask 2: Sentence classification
  - Does a sentence contain information about a past or ongoing event?
- Subtask 3: Event sentence coreference identification
  - Which event sentences (subtask 2) are about the same event?
- Subtask 4: Event extraction
  - What is the event trigger and its arguments?

We particularly focus on events that are in the scope of contentious politics and characterized by riots and social movements, i.e., the "repertoire of contention" (Giugni 1998, Tarrow 1994, Tilly 1984), which we name GLOCON Gold in our operationalization (Hürriyetoğlu et al. 2020a). The aim of the shared task is to detect and classify socio-political and crisis event information at document, sentence, cross-sentence, and token levels in a multilingual setting. The detailed description of the subtasks can be found in Hürriyetoğlu et al. (2019, 2020b). The data size for English is increased and data for Portuguese, Spanish, and Hindi are added in this edition.

Task 2: Fine-grained classification of Socio-political events

The objective of this task is to evaluate zero-shot learning event classification approaches to classify short text snippets reporting socio-political events with fine-grained event types using the Armed Conflict Location & Event Data Project (ACLED) event taxonomy, which consists of 25 event subtypes pertaining to political violence, demonstrations (rioting and protesting) and

selected non-violent, politically important events. One should keep in mind that the event definitions for task 1 and task 2 are not fully compatible.

Task 3: Discovering Black Lives Matter events in United States

This task is only an evaluation task where the participants of task 1 will have the possibility to evaluate their systems on reproducing a manually curated Black Lives Matter (BLM) related protest event list. Participants will use document collections provided by us to extract mainly place and date of the BLM events. The event definition applied for determining these events is the same as the one facilitated for task 1.

Data

There will be training and test data for each of the tasks and subtasks. Sample data, submission formats, scripts, baseline scores, application form, and any additional information will be shared on the dedicated online repository of the shared task: https://github.com/emerging-welfare/case-2021-shared-task. Copyright of the news articles is protected by sharing URLs and code (Docker image) for retrieving text of the articles using these URLs for subtask 1. In all other tasks and subtasks only relevant portions of the articles such as only event sentences will be utilized.

Training Data

Task 1:

This edition of the task 1 extends the data in English and includes training and test data in Spanish, and Portuguese. The format and approximate dataset sizes for each task will be comparable to the previous editions of the subtasks. However, the Spanish and Portuguese training data for subtasks 3 and 4 will be relatively less.

Task 2:

For the training purposes one will use a relatively large human-created/coded data set of event type-labeled short text snippets (circa 600K event records) extracted and curated from ACLED event database. The training data for this task is the "ACLED-III" event dataset described in Piskorsky et al. (2020) and available under http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/LANGUAGE-TECHNOLOGY/2020_annotated_event_dataset/Folds/.

Each line of the file in the corpus consists of three tab-delimited elements, namely: (a) text snippet reporting an event, (b) event main type, and (c) event subtype. In this subtask the focus is on the classification of events represented by the text snippets using one of the 25 subtypes (single-label classification problem).

Task 3:

There will not be any additional training data for the task 3. The systems developed for task 1 or task 2 should be used to process the test data that will be provided to the participants.

Test data

Task 1:

Test data for subtasks 1-4 will be in the formats described in Hürriyetoğlu et al. (2019, 2020b) and %25 of the training data, which is 80/20 split of the original data. There will be test data in English, Portuguese, and Spanish for all subtasks. Data in Hindi language will be available only for evaluation of the multilingual models for the subtask 1.

Task 2:

Test data for this task will be around 1,000 text snippets from news, web pages reporting socio-political events and artificially created event descriptions labelled using the ACLED event taxonomy, (not from ACLED). The registered participants will be provided a single file, where each line consists of three tab-separated elements, i.e., an ID (integer), followed by a text snippet reporting an event. The system response files should have per event a line with the event ID and an event label separated by a tab.

Task 3:

The test set for task 3 will consists of two separate and independent datasets that are a tweet dataset (tweet IDs) by Giorgi et al. (2020) and a list of URLs (or document IDs in the target news archive) to news articles. The code that can be used to access to this data will be provided by the organizers of the shared task.

Evaluation plan

Evaluation is carried out on the system responses returned by the participants on the test data for each task. The evaluations will be performed on Codalab (https://codalab.org/). Each team will be allowed to submit multiple valid system responses for each task or subtask. The ranking will be based on the best result of a team. The evaluation metrics for each task are provided below.

Task 1:

F1-macro will be calculated on the predictions on the test data for the subtasks 1, 2, and 4. We use conll-03 evaluation script for subtask 4. The subtask 3 will be evaluated using Adjusted Rand Score for the test data in each language. There will be a separate evaluation for each subtask in Task 1 using the test data for each separate language, which are English, Portuguese, Spanish, and Hindi.

Task 2:

The systems will be evaluated mainly using: Precision, Recall, and Micro and Macro F-1 metrics, where the last two are the most important ones.

Task 3:

The evaluation data will be a list of protest events pertaining to Black Lives Matter. Each event record should include information such as place and time of a single event. Spatio-temporal correlation between the manually curated event list and the submissions will be calculated to

determine the score for each submission, using an adaptation of the method used in Hammond and Weidman (2014) and applied for analysis of the dynamics of conflicts (Zavarella et al. 2020).

Participation

You can participate either individually or as a team. In any case, you should provide us with a list consisting of:

- Team name
- A contact person
- Contact email
- A list of the team members.

For the tasks 1 and 3, all members of a team should complete, sign, and send the application form, which can be found on the shared task repository with the name "CASE2021-Socio-political-and-Crisis-Events-Shared-Task-Individual-Application-Form.pdf", to Ali Hürriyetoğlu (ahurriyetoglu@ku.edu.tr).

For the task 2, there is no need to sign the application form. In order to participate and register for this task the aforementioned team details should be sent via email to case2021.task.finegrained@gmail.com.

Participation requests must be completed by the registration deadline, that is April 8.

Publication

Participants in the Shared Task are expected to submit a paper to the CASE 2021 workshop co-located with ACL-IJCNLP 2021 (https://emw.ku.edu.tr/case-2021/). Submitting a paper is not mandatory for participating in the Shared Task. Papers must follow the CASE 2021 workshop submission instructions (ACL 2021 style template: https://2021.aclweb.org/calls/papers) and will undergo regular peer review. Their acceptance will not depend on the results obtained in the shared task, but on the quality of the paper. Authors of accepted papers will be informed about the evaluation results of their systems prior to the paper submission deadline (see the important dates).

Contact

Please reach us using the following e-mail address for anything you may think we can support you: Ali Hürriyetoğlu, ahurriyetoglu@ku.edu.tr (Task 1 and Task 3 and any other issue), Jakub Piskorski case2021.task.finegrained@gmail.com (Task 2), Salvatore Giorgi, sgiorgi@sas.upenn.edu (Task 3, collecting an on the ground events list and using the tweet collection). The Github repo of the shared task (https://github.com/emerging-welfare/case-2021-shared-task will be updated regularly.

Important dates

Shared task announcement: February 23

Release of training data for Task 1: March 1, 2021, Task 2: already available
Registration deadline: 8 April 2021
Release of test data for all tasks to registered participants: 23 April 2021,
Submission of system responses: 26 April 2021 (12:00 CET)
Results announced to participants: 28 April 2021
Shared Task Papers Due: 10 May 2021
Notification of Acceptance: 28 May 2021
Camera-ready papers due: 7 June 2021
CASE 2021 Workshop (presentation of the ST results): 5-6 August 2021

All deadlines are 23:59 AoE (anywhere on Earth) and in the year 2021, unless otherwise stated above.

References

Giorgi, S., Guntuku, S. C., Rahman, M., Himelein-Wachowiak, M., Kwarteng, A., & Curtis, B. (2020). Twitter corpus of the# blacklivesmatter movement and counter protests: 2013 to 2020. arXiv preprint arXiv:2009.00596. URL: https://arxiv.org/abs/2009.00596, Dataset: https://zenodo.org/record/4056563, GitHub: https://github.com/sjgiorgi/blm_twitter_corpus

Giugni, Marco G. (1998). Was It Worth the Effort? The Outcomes and Consequences of Social Movements. Annual Review of Sociology 24 (January): 371–93. 1998. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev.soc.24.1.371

Hammond, J., & Weidmann, N. B. (2014). Using machine-coded event data for the micro-level study of political violence. Research & Politics, 1 (2). URL: https://journals.sagepub.com/doi/full/10.1177/2053168014539924

Hürriyetoğlu A., Yörük E., Yüret D., Mutlu O., Yoltar Ç., Duruşan F., Gürel B. (2020a). Cross-context News Corpus for Protest Events related Knowledge Base Construction. In the Proceedings of Automatic Knowledge Base Construction Conference. URL: https://doi.org/doi:10.24432/C5D59R

Hürriyetoğlu A., Zavarella V., Tanev H., Yörük E., Safaya A., and Mutlu O. (2020b) Automated extraction of socio-political events from news (AESPEN): Workshop and shared task report. In Proceedings of the Workshop on Automated Extraction of Socio-political Events from News, pages 1{6, Marseille, France, May 2020. European Language Resources Association (ELRA). ISBN 979-10-95546-50-4. URL: https://www.aclweb.org/anthology/2020.aespen-1.1.

Hürriyetoğlu A., Yörük E., Yüret D., Yoltar Ç., , Gürel B., Duruşan F., Mutlu O., and Akdemir A. (2019) Overview of Clef 2019 Lab Protestnews: Extracting Protests from News in a Cross-context Setting. In Proceedings of the Conference Experimental IR Meets Multilinguality, Multimodality, and Interaction, pages 425{432, Cham, 2019b. Springer International Publishing. ISBN 978-3-030-28577-7. URL: http://ceur-ws.org/Vol-2380/paper_249.pdf

Piskorski, J., Haneczok, J., & Jacquet, G. (2020). New Benchmark Corpus and Models for Fine-grained Event Classification: To BERT or not to BERT?. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 6663-6678). URL: https://www.aclweb.org/anthology/2020.coling-main.584.pdf

Tarrow, S. (1994). Power in Movement: Social Movements, Collective Action and Politics. Cambridge, UK: Cambridge University Press. URL: https://doi.org/10.1017/CBO9780511813245

Tilly, C. (1984). Big Structures, Large Processes, Huge Comparisons. New York: Russell Sage Foundation. URL: https://www.jstor.org/stable/10.7758/9781610447720

Zavarella, V., Piskorski, J., Ignat, C., Tanev, H., & Atkinson, M. (2020). Mastering the Media Hype: Methods for Deduplication of Conflict Events from News Reports. In Proceedings of the Workshop Proceedings of the Artificial Intelligence for Narratives (AI4Narratives). URL: http://ceur-ws.org/Vol-2794/paper6.pdf