TASK 3: DISCOVERING BLACK LIVES MATTER EVENTS IN UNITED STATES

1. Task Definition

The goal of this task is to evaluate the performance of automatic event detection systems on modeling the spatial and temporal pattern of a social protest movement. Namely, the participants will have the possibility to evaluate their systems on reproducing a manually curated Black Lives Matter (BLM)-related protest event dataset, by detecting BLM event reports, enriched with location and date attributes, from a Twitter collection and a news corpus we provide. The event definition applied for coding the reference event dataset is the same as the one adopted for Task 1. Participants may utilize any other data source to improve performance of their submissions. System performance will be evaluated by computing correlation coefficients on event counts spatially aggregated on uniform grid geographical cells.

2. Training Data

No training data are provided for this Task. The data utilized for Task 1 and Task 2, such as e.g. the one from Hürriyetoğlu, A. et al. (2020b) or any additional data can be used to build systems/models that can detect Black Lives Matter (BLM) events in tweets and news articles.

3. Test Data

The meta-data that can be obtained from Twitter or New York Times can be used to improve the results. For instance the location fields of a tweet.

4. Evaluation

In order to evaluate the ability of automatic event-coders to reproduce the gold standard BLM dataset, we adapt here two correlation methods originally used in micro-level analysis of political violence by Hammond and Weidmann (2014) (see also Zavarella et al. 2020).

First, we aggregate event counts in both system response and gold standard on uniform grid geographical cells, by using PRIO-GRID, a vector grid network with a resolution of 0.5 x 0.5 decimal degrees (~55 km), covering all terrestrial areas of the world (Tollefsen et al.,2012). Then, for each day in the time range of the reference dataset, we compute the total number of

grid cells experiencing one or more protest event, and we apply a number of correlation coefficient/error measures to these values, namely: **Root Mean Squared Error**, **Pearson coefficient** and **Spearman's rank correlation coefficient**.

In order to have a more fine-grained representation of the protest movement intensity, we also compute the same correlation measures above between absolute event counts per day-cell units from system response and gold standard dataset.

For each input text corpus in 3., each participant may submit up to 3 different system responses. Each system response will consist of a csv file with the following naming pattern:

"submission.<team-name>.<corpus>.<response-number>.csv"

where corpus is either "twitter" or "nyt".

For instance:

"submission.MyTeam.nyt.3.csv" for the 3rd submission of team "MyTeam" on the New York Times corpus.

Each system response file will have one line per event, where each line will have the following format:

<id>,<City>,<Region>,<Country>,<Date>

where:

<id> is a numerical event identifier

<City>,<Region>,<Country> are canonical English names of the City,State/Region and Country, respectively, of the detected event location. For the purpose of evaluation, the String representation of the event location will be automatically converted into a pair of geographical coordinates by using the Nominatim search API (https://nominatim.org/release-docs/develop/api/Search/). While only the <*country*> attribute is mandatory, systems are expected to assign a description of the event location at the finest grained level possible, as otherwise geographical coordinate conversion from Nominatim will place the event at the geographical centroid of the polygon of the assigned administrative unit, which may penalize the correlation score on Prio-GRID cell aggregation.

<Date> is the assigned date of the event in the format YYYY-MM-DD

A sample system response file line:

0,Pelham,New York,USA,2020-06-15

A sample system output file can be downloaded from the Task repo at:

https://github.com/emerging-welfare/case-2021-shared-task/blob/0536411b1d00a83da89ec7034d7a8c28cfb8b5c7/task3/submission.myteam.nyt.3.csv

Note: for the system runs on the tweet corpus, in order to compensate for the lack of contextual information and mitigate ambiguity of location names, we allow participants to provide underspecified location descriptions of type <city>,<null>,<null> like e.g.: *"Pelham,,".* For the set of target locations in the gold standard, this will be normalized to the correct, fully specified form: (e.g. "*Pelham,New York,USA"*). However, in order to reward systems that are able to disambiguate location information, we will introduce noise in the system responses, in the form of extra event rows placed at the uncorrect locations of the homonym place names. The noise penalization factor will be proportional to the overall amount of ambiguity present in the system responses (number of underspecified event locations times number of actual homonym places with that name).

5. References

Jesse Hammond and Nils B Weidmann. Using machine-coded event data for the micro-level study of political violence. Research & Politics, 1(2):2053168014539924, 2014.

Hürriyetoğlu, A., Yörük, E., Yüret, D., Mutlu, O., Yoltar, Ç., Duruşan, F., & Gürel, B. (2020b). Cross-context news corpus for protest events related knowledge base construction. arXiv preprint arXiv:2008.00351. In Automated Knowledge Base Construction (AKBC). URL: https://www.akbc.ws/2020/papers/7NZkNhLCjp

Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. Journal of Peace Research, 49(2), 363-374

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: A. Jorge, R. Campos, A. Jatowt, A. Aizawa (eds.): Proceedings of the first AI4Narratives Workshop, Yokohama, Japan, January 2021, published at http://ceur-ws.org