Review 1

General Comments This is a review paper summarizing at a high level currently available datasets and approaches that tackle crisis analytics on Twitter data, as well as highlights some challenges. The paper is overall well written and should be understandable by an in-expert audience. It may be valuable as an entry point for new researchers looking to work in the area of crisis informatics (although there are already good resources available in this area (e.g. the Big Crisis Data book). On the other hand, the paper is quite shallow in terms of detail in all aspects and so only acts as a guidepost for further reading on the subject. There are no individual experiments by the authors provided in the paper.

Detailed comments The core limitation of this work from my reading is that it tries to summarize too many areas of the field of crisis informatics and so currently does not provide enough detail on any one to provide significant insights that add value over the individual works. For example, Section 2 provides a brief summary of some of the definitions that different groups have used for analysing crisis content, but don't go into detail on what works are compatible with one another, or indeed provide information
on the definition of 'Informative' regards to who is used in each work. As a result it is not clear to someone just entering the field what they should read.

- ➔ We have expanded this section to give recommendations for handling this issue in future research. Newer publications have shifted to a more user-centric definition (actionability) and we show this as a path forward both in this section and in the rest of the paper.

Similarly, Section 3 highlights some datasets used in crisis informatics, but Table 2 only lists tweet counts, not the volume labelled and what was labelled or for what task. Which datasets are complementary? Which are easy to work with? What datasets do the authors recommend researchers use?

- ➔ Labeling information has been added to the table. At the end of the section, we have added a comparison of what data sets are useful for what purposes.

The second question I have for this overview is whether COVID-19 datasets belong in this study or whether they should be considered separately. COVID is quite a different task compared to natural or manmade disasters, as there typically is not a strong timeliness component to related information needs. Hence, the definition of what is informative for pandemics and the associated target user groups are very different. I would recommend at least adding some discussion in Section 2 on this point.

- ➔ COVID-19 data sets have been removed from the overview, and COVID-19 data analysis is now mentioned as an interesting new research task. We think it would be interesting to get started on a separate review paper about COVID-19 data sets and approaches in a year or so.

Third, I would recommend structuring the discussion on the machine learning aspects along the lines of what task is being investigated, inputs, features and models. Indeed, it would be valuable to get some idea of how many works use each different approach, as well as get some data on the prevalence of different feature and text representation approaches used in the different works and critically, what patterns emerge on what works.

- ➔ Section has been expanded with more ML approaches (particularly non-neural network ones) and a table comparing methods, features, tasks, and used data. Subsections have been

restructured to better compare used feature and ML methods.

Other notes: - Table 2 should highlight the differences between labelled and un-labelled tweets - I believe the statistics for the TREC-IS data in particular is out-of-date, see the ISCRAM 2020 paper http://trecis.org/2020/ISCRAM_2020_TREC_IS.pdf

➔ Has been updated, table now includes statistics of labeled vs. unlabeled tweets

Review 2

General comments The review paper summarizes research on the automatic processing of social media messages in crisis situations. The paper outlines varying concepts of information, relevance, available datasets, filtering approaches and associated challenges. Overall, and echoing the concern of the previous reviewer, the paper can serve as an introduction to this research but lacks the comprehensive detail or specific focus that would recommend it to experts. The authors might consider approaching the review from the perspective of one of the challenges highlighted in section five with the goal of providing an overview of approaches addressing this challenge and revealing new directions for research.

➔ We have chosen the challenge of varying definitions of relatedness/relevance/importance as this is also the start of the paper. The paper now provides possible solutions in each section.

Specific comments The different understandings of information relevance between research focusing on technical challenges related to processing social media messages and research focusing on the information requirements of end users, i.e., emergency responders, deserves further attention. As the authors point out, the former includes different concepts of information relevance ("related", "relevant", and "informative") that are not often compared or resolved in research addressing technical challenges. Moreover, compared to research on end user requirements, these definitions are considered much too coarse grained. When employed in the design of social media filtering tools, these designs may provide information that is somehow related to a crisis but typically unrelated to responders' information needs which not only vary by role and crisis, but by dynamic contextual factors such as the availability of information from existing, traditional information sources. Recent research has considered these challenges by turning to the concept of "actionability" to describe information relevance from the end user perspective of emergency responders (Kropczynski et al., 2018; McCreadie et a., 2020; Zade et al., 2018).

➔ Has been included and is now a focal point of the paper as described above.

As the paper provides a cursory overview of available datasets, filtering approaches and associated challenges, greater detail might be included in each.

➔ Data and approach sections have been expanded and the comparison has been made more detailed

As described above, selecting a particular research challenge as a way to focus the review and discuss in greater detail the contributions of these studies with respect to this challenge would offer readers more insight into the research space.
There are additional datasets that might be included, such as listed in Palen et al. (2020) and Grace (2020).

➔ Grace data set and additional literature have been included (including Palen et al.)