Review 1

The paper describes how information extracted from the micro-blogging platform Twitter can serve Crisis Management during a disaster rather than finding actionable tweets.

The structure of the paper is clear, and the data are exhaustively detailed. Therefore the presentation quality is good. I have significant doubts about the scientific significance and the quality of the outcomes.

Scientific significance

Contribution to the research community:

a-The vast majority of SM research in DRM uses Tweets as they are the 'de-facto' only open and free source of SM data

b-The authors start the paper in the quest for actionability but then end up describing mostly research works used for classifying relevance of tweets (4.2 and 4.3). Neither the datasets described classified tweets as 'actionable' but rather as informative or relevant.

> ➔ This difficulty is pointed out in the paper; previous research mainly focuses on "relevance" or "informativeness", not on actionability. We have now clarified that these works form a basis for adaptation to actionability.

As it is now, the paper offers an update of similar works already published (HAL, Reuter work to cite two). The novel information is the in-depth description of the models and technologies used for automated classification of tweets. The authors should leverage on the richness of the examples offered to better present the state of the art of several lines of research according to crisis management needs.

Scientific quality

My concern is that the paper's goal is not clear (or the research question not posed correctly?). I cannot understand if the article wants to offer a path for future research on finding serviceable information or describing the state of the art of SM for Crisis Management.

In the first case, the paper misses a solid base of works for supporting a choice; in the second case, there is an unclear definition for 'actionable information'.

> ➔ In a sense, both. We present the current state of the art, but point out the unclear definition of what exactly approaches published so far are supposed to classify. As suggested by the other reviewers, we then introduce the concept of actionability in the sense of letting the user define what is relevant to them/their use case. On methods, we discuss first what is already there, and second how actionability can be introduced. The paper has been updated to clarify our approach to the problem and the progression from existing works to methods for actionability. The definition of actionability is also stated more clearly now.

A tentative definition is quickly described in the framing of the problem, but then it is not followed in the next chapters of the article (4.2 and 4.3). The authors describe models for classification for situational awareness in the machine learning approaches justifying it as possible pre-selection for further analysis for finding actionable information. That's perhaps confusing. There should be a more

precise definition of classes of actions/decisions and clear links between each data/algorithm/technology described and such classes.

➔ We explicitly do not focus on specific use cases or events because we are more interested in making systems usable for various ones, even those that may come up in the future and are not yet predictable (as we are now seeing in the pandemic). We have now included this motivation in the paper.

Another concern is that if the paper aims to offer a starting point for further reading, I see the risk of highlighting only some branches of the research from the community of researchers studying the intersection of Crisis Informatics and Social Media, which is significant and growing (focus seems only on the authors that published data in the datasets section/par).

Some missing ref for instance (easily found on Arxiv or Google Scholar):

Purohit H. efforts for serviceability of information is missing (Purohit, Peterson).

H2020 funded Projects such as E2MC or I-REACT are not cited while they could be listed in the crowdsourcing approach (both produced publications).

Castillo is cited but not for his last works where Convnet is used for multilingual annotation (Lorini et al., ISCRAM 19 paper award)

Again no mention of De Bruijn or Brouwer works about the detection of impact (damage assessment) very relevant together with the cited Albuquerque research about mixing authoritative and social media data.

Poblete research in Disaster Management focuses on impact detection and multilingual text.

➔ References have been included

I also think leaving out the analysis of research on Images (growing area of study) from Social Media is also a missed opportunity to describe a growing source of informative data.

➔ We agree that this is a very interesting new development, but have not included it in the paper in order to maintain a text-centric focus. If there is interest in this, we are open to expanding the paper in this direction.

Suggestion for revision:

As it is now the work has a good depth, but I am afraid is mixing different researches and experiments. It would be better, in my opinion, to change the aim of the paper as to describe how Crisis Informatics research try to extract from SM information that can be 'fit-for-purpose' in the context of Crisis Management and then divide the work into purposes (finding requests for action, situational awareness, Impact assessment, ...).

Restructuring the paper according to 'purposes' or 'needs' would simplify the 'labelling' of the described works and streamline the narrative (extending the references of course).

➔ See above; we would prefer not to limit ourselves to certain use cases. We hope the new version of the paper conveys more clearly what we mean by focusing on actionability, and how this can be a path forward for many new use cases (i.e. the mentioned 'fit-for-purpose' data retrieval) in the future.
We thank the reviewer for their helpful comments and suggested references!