First, we sincerely thank you for your insightful and constructive comments. In what follows, we respond point-by-point to the comments. For comments that consist of multiple questions, we add numerical indices such as 1) and 2) in the comments to help identify the corresponding responses. Each comment is in a normal format and listed first, followed by our response in blue and the corresponding revised texts in the manuscript in green.

1. NOVELTY OF THE STUDY. The research on using Weibo data for early warning of water logging early warning seems to be an interesting idea. However, *1)* as this analysis is presented only for a specific case it remains unclear which scientific universal results and conclusions can be drawn, which would be interesting for the international scientific community. For the reader, it is currently not clear why someone outside of the study area should be interested in the results. In the discussion, *2)* the new universal insights should be explained and to which settings these are transferable. This would also *3)* require a more detailed description of the settings in your case study and an in-depth discussion of the algorithms and assumptions used to support early warning for waterlogging situations. Currently, the results are not interpreted or discussed sufficiently in this context. *4)* For instance, time aspects as regards processing time and warning lead times achieved are not mentioned.

1) This question is very enlightening for us. Both universality and novelty are important in model building.

   As for universality, a pioneer study using social media data for disaster detection and early warning is Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, Sakaki et al. In this paper, promote and verify two assumptions: 1. Each Twitter user is regarded as a sensor. A sensor detects a target event and makes a report probabilistically; 2. Each tweet is associated with a time and location, which is a set of latitude and longitude. This idea quickly spreads to other natural disasters. This paper applies this idea to the real-time urban waterlogging disaster detection and verifies the feasibility in two cases (heavy rainstorms in Beijing on August 12, 2020 (case A) and July 18, 2021 (case B)), over 20,000 microblog texts. We also investigate studies on the open data in other natural disasters and list their data sets below.

Paper	Dataset
Urban waterlogging risk assessment based on internet open data: A case study in China, https://doi.org/10.1016/j.habitatint.2017.11.013	122 reports in one case (Xiamen)
Real-time identification of urban rainstorm waterlogging disasters based on Weibo big data, https://doi:10.1007/s11069-018-3427-4	3,261 microblogs in one case
Analyzing the Evolution of Rare Events via Social Media Data and k-means Clustering Algorithm, https://10.1109/ICNSC.2016.7479041	74,000 microblogs in one case
This paper	22,631 microblogs in two cases

In the future, we will expand our database for some more universal results. Now, as our idea is borrowed from other mature research, our main technical method, SVM, is also a classical classification method, and the data set is also sufficient among similar studies, we consider the results credible.

As for the novelty, apart from word vector, we introduce sentiment analysis and publisher features for the classification task. Then, we build an intensity dictionary for assessing the risk value in each microblog and label the potential waterlogging area on the map. Our research does not stop at the microblog filtering stage but proposes an early warning model that can be used in emergency management practice.

2) We have modified the following paragraph to explain the new universal insights and transferable settings.

The 1st paragraph in section 4.2:

Unlike earthquakes, a rainstorm is predictable. Thus, there are many types of information about the rainstorm disaster, for example, the warning news before the storm, the doubts about the forecast, and the rescue reports after the storm. Such information should be separated from the real-time rainstorm information, as they contribute little to the real-time disaster assessment and early warning. In this paper, we improved the extraction results of a rainstorm and waterlogging information in two aspects.

The 6th paragraph in section 4.2:

From the table above, we could get a higher accuracy within one disaster data group. In migratory validation, the accuracy would be slightly worse, but it could be further improved with a larger training set. In the experiments, it was especially difficult to classify short texts with less than 10 words. People usually used some objects to represent rainstorms and waterlogging, such as shoes or windows, which rarely appeared in our training set. This situation reduced the accuracy of the classification. Overall, it showed the feasibility of being applied in practice that we can further separate the real-time disaster information from all the disaster-related texts in such predictable disasters.

The 8th paragraph in section 4.2:

There are three types of publishers on the Weibo platform: official authentication, personal certification, and the general user. Previous studies have considered that information from three types of publishers is equally valuable (Na Xiao et al. (2018)), while we discuss their difference in timeliness in this part. We counted the number of microblogs on rainstorms and waterlogging posted by three types of users in each hour from 11 am on August 12 to 11 am on August 13. As the number of each publisher type was different, we calculated the percentage of posts in each hour for different types. The results are shown in figure 3.

The 9th paragraph in section 4.2:

There were two large-scale precipitations in this period. One with less rainfall lasted from 11 am to 2 pm. The other heavy precipitation lasted from 9 pm to early morning the next day. For the peak position, there was a lag of about 1 hour for official authentication users, which means that their microblogs may contain outdated information. For the peak height, the number of microblogs posted by the general users was positively correlated with precipitation. Therefore, when we used social network data to evaluate the degree of urban rainstorms and waterlogging disasters, much attention must be paid to the microblogs of general users.

The last paragraph in section 4.3:

We could conclude from figure 6 that the inverse distance function was better. Of the total nine waterlogging points, four were in the area with a value over 3, located in Zone 2, Zone 4, Zone 6, and Zone 7. Three points were in the local extreme area, located in Zone4, Zone 7, and Zone 8. What's more, in eight areas with a greater risk of waterlogging, four had waterlogging points, including Zone2, Zone 4, Zone 6, and Zone 7. This showed that the high-risk areas of waterlogging obtained from Weibo texts with location information were certainly accurate. We could also notice that some waterlogging points were located in the low-risk area. The first reason was the conditions for the formation of waterlogging were complex, in addition to precipitation, including topography, drainage systems, and other reasons as well. Another reason was the limit on the number of microblogs, which led to inaccurate assessments in marginal areas. The results show that social media users, as social sensors, don't just transmit yes or no information. Instead, we can dig out more detailed information such as intensity by quantifying keywords for emergency management.

Conclusion:

Before modification	After modification
This paper proposed a social network data processing method for an urban rainstorm and waterlogging disaster risk assessment and real-time detection. Based on the word vector, we could separate the microblogs with timely disaster information. Combining the classification results with the publisher feature and sentiment analysis, we could better understand the time and severity of the rainstorm and waterlogging disasters. Microblogs posted by general users can better represent the intensity and timing characteristics of precipitation and microblogs posted by personal certification users are also timely. Furthermore, we built an urban rainstorm and waterlogging disaster dictionary for real-time risk assessment and early warning. With microblogs with location information, we could generate a real-time waterlogging risk map for emergency management.	This paper proposed a social network data analyzing model for an urban rainstorm and waterlogging disaster's risk assessment and real-time detection. The novelties of this study lie in three main aspects. First, we screen microblogs with timely disaster situations from all the disaster-related microblogs. Unlike earthquakes, rainstorms are predictable. Therefore, many users may express their concerns or expectations on the rainstorm by microblogs which carry little useful timely disaster information. Based on the word and sentiment vector, we could further separate the microblogs with timely disaster information. The classification accuracy in the same case reached 71.73%, and 66.62% in different cases. Second, from the perspective of publisher features, general users publish the timeliest disaster

	information while there is usually a delay for personal and official authentication users' microblogs. This result shows that researchers should pay more attention to microblogs by general users when it comes to determine the starting time. Third, in the selected microblogs, we build an urban rainstorm and waterlogging disaster intensity dictionary for waterlogging risk assessment. By analyzing the disaster levels of different microblogs, we get a real-time risk assessment map by ArcGIS and inverse distance weighted interpolation for emergency management and early warning. The waterlogging spots summarized in the report verify the model's accuracy. Social media users, as social sensors, don't just transmit yes or no information. Instead, we can dig out more information such as intensity for emergency management.

3) We have added the following sentences for an in-depth discussion of the algorithms and assumptions.

   Penultimate paragraph in Section 1:

   Rainstorms are predictable, but there is still some error in time, range, and intensity. What's more, the secondary disaster followed by rainstorms, the urban waterlogging, can be hardly predicted. Based on the discussion above, the huge amount of Weibo users can be regarded as social sensors and there are time, location, and intensity information on ongoing disasters in their microblogs (Sakaki et al. (2010)). By extracting appropriate features, we can select these disaster-related microblogs with classification algorithms such as SVM (Nair et al. (2017)). These microblogs tell us the intensity of precipitation and the depth of waterlogging in different places at different times, which can help us identify spots of potential waterlogging.

4) We quite agree that warning lead time is an important parameter in the early warning

model. However, the exact formation time of waterlogging is unknown. All we know from the report the next day is that there was flooding somewhere. Therefore, in this paper, we cannot give the exact lead time. The results of this model are more like risk assessment and helping emergency management departments to identify and patrol key areas.

Processing time is added in the following sentences.

The last paragraph in Section 4.3:

…Compared to deep learning, the calculation process of SVM and inverse distance weighted interpolation is not complicated, which ensures the timeliness of the model's results…

2. TRANSPARENCY AND REPRODUCIBILITY. The description of the proposed method is quite technical but *1)* it lacks motivation why these algorithms have been selected. *2)* Many assumptions lack justification and their implications remain unclear. *3)* It is acknowledged that you make your code openly available but also the data sources should be clearly referenced and accessibility should be given.

1) This comment can help us improve the structure of the paper. We have added the following sentences to illustrate why we choose these algorithms.

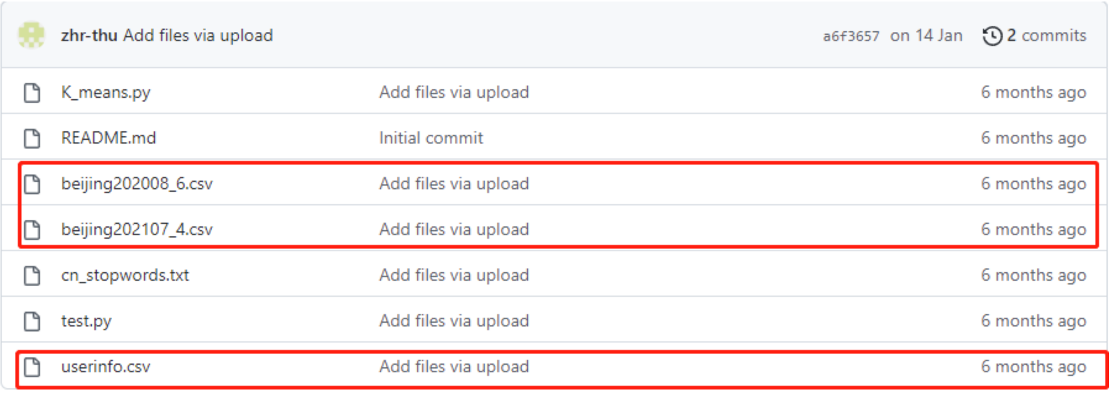
Before modification	**After modification**
The k-means clustering analysis is a simple and commonly used clustering algorithm based on the squared error criterion which is over 50 years old (MacQueen (1967)).	The k-means clustering analysis is a simple and commonly used clustering algorithm based on the squared error criterion which is over 50 years old (MacQueen (1967)). The k-means algorithm can process the original data with weak regularity and form the initial classification with a certain regularity
SVM is a linear classifier defined in feature space with the largest interval, which distinguishes it from perceptron. With the help of different kernel functions, it can also handle the nonlinear problem.	SVM is a linear classifier defined in feature space with the largest interval, which distinguishes it from perceptron. With the help of different kernel functions, it can also handle the nonlinear problem. In the case of low dimensional or sparse features, SVM performs as well as a deep learning network and has a faster processing speed.
In this paper, we applied the Inverse Distance	In this paper, we applied the Inverse Distance

Weighted (IDW) interpolation and the local polynomial interpolation (Caruso et al. (1998)) for comparison.	Weighted (IDW) interpolation and the local polynomial interpolation (Caruso et al. (1998)) for comparison. IDW interpolation implements a basic law of geography; i.e. things that are close to one another are more alike than things that are far apart. All the distances have the same power, which is 2 in this paper. The polynomial interpolation assumes that every point on the surface conforms to some polynomial formula. Global methods use all the known values to estimate an unknown one, while in local methods only a specified number of nearest neighbors are used. We choose the local method as the area in our research contains 6 districts and the terrain may be different. To avoid the violent fluctuation of value caused by over-fitting, we choose power 1 in this method.

2) We have added the following sentences to supplement the assumptions.

Before modification	After modification
There is an assumption that if a microblog is reposted, it is more likely to carry outdated information, which cannot be used for real-time disaster assessment, so we eliminate all the repost text.	There is an assumption that if a microblog is reposted, it is more likely to carry outdated information. There is always a time lag between the release time of the microblog and the disaster situation described. For the original microblogs, we can infer the time lag from the words such as now, just, and so on. However, in the case of reposted microblogs, there is an extra time lag for the second user to notice the microblog, so we eliminate all the reposted microblogs.
Adding the three degrees together, we gave the waterlogging risk value in different blocks. We assumed that areas with a value greater than 3 had a high risk of waterlogging.	Adding the three degrees together, we gave the waterlogging risk value in different blocks. As there were three indexes and a value of 1 represents a moderate risk in each index, we assumed that areas with a value greater than 3 had a high risk of waterlogging.

3) The data set is crawled by ourselves and placed in the same open link with the code.


zhr-thu Add files via upload a6f3657 on 14 Jan 2 commits

K_means.py	Add files via upload	6 months ago
README.md	Initial commit	6 months ago
beijing202008_6.csv	Add files via upload	6 months ago
beijing202107_4.csv	Add files via upload	6 months ago
cn_stopwords.txt	Add files via upload	6 months ago
test.py	Add files via upload	6 months ago
userinfo.csv	Add files via upload	6 months ago

3. FIGURES AND TABLES Several figure captions and table captions are exceptionally short and are not self-standing.

We have reviewed all the figure captions and table captions and made the following changes.

Before modification	**After modification**
Table 1. Examples of positive and negative Weibo texts	Table 1. Examples of positive (with timely disaster situation) and negative (without timely disaster situation) Weibo texts
Table 3. The classification accuracy (unit: %)	Table 3. Comparison of the macro-average indicators in two situations (unit: %)
Table 4. The dictionary of waterlogging degree	Table 4. The intensity dictionary for assessing waterlogging risk
Figure 2. The evaluation for different number of categories the elbow and silhouette coefficient method	Figure 2. Evaluation for different number of categories in k-means by the elbow law and silhouette coefficient method
Figure 3. The percentage of posts for different types in each hour.	Figure 3. Hourly microblogs post ratio to the total for three types of publishers