Research on Extracting Emergency Information from Crowd Sourcing Data

Weijun Ma (1,2), Deqin Liu (1), Hongliang Guan (2), and Yuqun Li (3)

(1) Chinese Academy of Surveying and Mapping, Beijing, China (weijunma@casm.ac.cn), (2) Capital Normal University, Beijing 100048, China., (3) Peace Map Co.,Ltd,Beijing 100083, China.

1. Background

With the development of Internet, especially the Mobile Internet, the crowd sourcing data increased sharply. There is a large amount of information hided in the crowd sourcing data. The emergency information is the most useful one, such the exact location of an earthquake, the photos of the emergency etc. A method for getting the emergency information from the crowd sourcing data is provided in this paper.

2. Method

There are five main steps to get the emergency information and to use them:

1)Building the database of key words and the sample database of the emergency

The emergency is divided into four types: the Natural disasters, the accident, the Public health and the social security. The key words database of the four types was established. And we collected about 1000 samples of four kind of emergency to serve as the sample database of the emergency.

2) Finding the source of the emergency information from the internet based on the web crawler technology.

3) Getting the emergency information from the crowd sourcing data based on the Natural Language processing technology.

4) Getting the key information of emergency, such as the time, location and type of the emergency etc., and classifying the emergency into one of the four types according to the characteristic. A method based on The Bayesian Classifier is used in this paper.

5) Spatializing the emergency information based on the geocoding technology, then integrated with the population data and the others social economic data to evaluate the emergency damage.

3. Experiment

To test the method above, an experiment about how to get information from the News web site is descried in this paper.

4. Conclusion

By using the method provided in this paper, we can find the emergency information hided in the crowd sourcing data effectively, the method is valid.