An Integrated Method for Assessing the Text Content Quality of Volunteered Geographic Information in Disaster Management

Kuo-Chih Hung, University of Melbourne, Melbourne, Australia
Mohsen Kalantari, University of Melbourne, Melbourne, Australia
Abbas Rajabifard, University of Melbourne, Melbourne, Australia

ABSTRACT
Volunteered geographic information (VGI) has the potential to provide much-needed information for emergency management stakeholders. However, stakeholders often lack scalability to identify useful and high-quality text content from the often-overwhelming amount of information. To solve this problem, most studies have concentrated on using text-related features in supervised learning models to classify text contents. This article proposes an assumption that the geographic attributes of VGI can be integrated into the model as features for enhancing the model’s performance. To evaluate this assumption, the authors developed a case study based on VGI collected from two flooding events in Brisbane. They validated the accuracy of associated geographic coordinates and defined the geographic features relevant to the flood phenomenon. From their experiments, model based on this integrated method can have better performance in comparison with the model trained from the text-related features. The results suggest great potential for using the integrated method to harvest useful VGI for the needs of disaster management.

KEYWORDS
Volunteered Geographic Information (VGI), Disaster Management, Information Quality, Supervised Learning, Geographic Information System (GIS), Floods

INTRODUCTION
The growth of various web 2.0 applications (e.g. forum, interactive mapping platforms, and social media) provides a new approach to create and disseminate user-generated geographic content over the Internet. Such content is termed as Volunteered Geographic Information in the domain of Geographic Information Science (VGI, Goodchild, 2007). In this research, we focused on a subtype of VGI, text-based VGI, which consists of short texts with geographic attributes.

Text-based VGI recently played a critical role in disaster management. It has the potential to provide crucial information and enhance situational awareness, particularly in chaotic situations of response of extreme events. For example, in several extreme events such as the 2010 Haiti Earthquake and the 2011 Queensland Floods, an emergency reporting platform “Ushahidi” was used by the volunteers to gather disaster information. Affected residents can report the presence and extent of incidents; the volunteers can aggregate information from multiple sources on the platform (McDougall, 2009).
Ubiquitous social media is another important source of text-based VGI. Emergency Management (EM) stakeholders have used the social media stream in the communication, early identification of incidents, and emergency planning (Arklay, 2012; Shih, Han, & Carroll, 2015). These cases reflect that information access in mass emergencies has shifted to non-conventional models on the Internet. The use of VGI enhances the understanding of what is happening in existing conditions and how it could change over time.

Although text-based VGI can contribute significantly to disaster management, there is a major challenge in its use. In the context of extreme events, a huge number of VGI instances are available from relevant web applications. Some may provide informative content. Others may relate to an event but not provide any value to EM stakeholders. The text content quality of each information instance is hard to be assessed efficiently by manual efforts. In the literature, an automatic method for VGI text content quality assessment is to use the supervised learning model (e.g. Imran, Elbassuoni, Castillo, Diaz, & Meier, 2013; Imran, Mitra, & Srivastava, 2016). The supervised learning model can classify data to generate quality indicators. Currently, most previous work has sought a standard text classification approach which used only text-related features such as n-grams in the learning model. In this research, we propose an assumption that the geographic attributes of VGI (i.e. geographic features) can be used for modelling and can enhance the performance of learning models. According to Tobler’s first law of geography, “everything is related to everything else, but near things are more related than distant things (Tobler, 1970).” We assume that if a VGI instance is close to the flood-affected areas or has an informative neighbour spatially nearby, it is more likely to provide informative content as well. This geographic assumption has been discussed in a few case studies (e.g. Albuquerque, Herfort, Brenning, & Zipf, 2015; Craglia, Ostermann, & Spinsanti, 2012) but not yet evaluated in a small-scale study area (e.g. within a city) and via a supervised learning approach. Therefore, our research question is: within a small-scale study area, can geographic features be used for the development of a supervised learning model, and provide better performances than models without geographic features?

To evaluate our assumption, we propose an integrated method for VGI text content quality assessment. This integrated method combines text-related features and geographic features to train a supervised learning model. We tested the performance of this method in a case study using data collected from two previous flooding events in Brisbane, Australia. The remainder of this paper is structured as follows. Section 2 provides a literature review about the development of text-based VGI quality assessment methods. In Section 3, we present the research objectives and methodology. Section 4 describes the case study including the models and evaluations. Section 5 discusses the findings and limitations of our experiment and concludes this article.

LITERATURE REVIEW: QUALITY ASSESSMENT METHODS OF TEXT-BASED VGI

This section intends to present the state-of-the-art research in text-based VGI quality assessment methods, particularly in the context of disaster management. Before going into more detail, we first clarify the definition of quality indicators in this research.

Definition of Quality Indicators

Information quality is defined as the extent to which information can be used by those who apply it (Wang & Strong, 1996). As text-based VGI is generated from lay people, it is often insufficiently structured and possesses uncertain quality. The target of VGI quality assessment method can be classified into two types: the cluster of instances and the individual instances. With respect to the cluster of VGI instances, the quality assessment process aims to find credible information about major hazards and incidents (e.g. Bahir & Peled, 2015). Some case studies have developed the associated methods such as aggregation (De Longueville, Annoni, Schade, Ostlaender, & Whitmore, 2010;
Towards a Collaborative Disaster Management Service Framework Using Mobile and Web Applications: A Survey and Future Scope
[www.igi-global.com/chapter/towards-a-collaborative-disaster-management-service-framework-using-mobile-and-web-applications/207579?camid=4v1a](www.igi-global.com/chapter/towards-a-collaborative-disaster-management-service-framework-using-mobile-and-web-applications/207579?camid=4v1a)

Identify Disaster as a Possibility
(2000). *A Primer for Disaster Recovery Planning in an IT Environment* (pp. 4-8).
[www.igi-global.com/chapter/identify-disaster-possibility/119782?camid=4v1a](www.igi-global.com/chapter/identify-disaster-possibility/119782?camid=4v1a)