Quality of User-Reported Data in Health Virtual Community

TREO Talk Paper

Hui Wang
Dept. of Information Technology
Radford University
hwang26@radford.edu

Yabing Jiang
Information Systems & Operations Management
Florida Gulf Coast University
yjiang@fgcu.edu

Abstract

Virtual communities play an increasingly important role in healthcare services. In some health virtual communities (HVC), user-reported data are used in medical research and to support healthcare operations. However, the quality of user-reported data presents a major challenge. We are interested in investigating the data quality issues in HVC and possible causes and improvement.

A recent survey by PwC's Health Research Institute finds that one out of three consumers surveyed uses social media for health-related matters (PwC HRI Social Media Consumer Survey 2012). HVCs, such as PatientsLikeMe, connect individuals with similar medical conditions and health concerns. Participants share information, and gain and give social support. Users report data such as personal profile, symptom, diagnosis, treatment, disease progression, and ratings of healthcare providers. These data are used for tracking and profiling (Barrett et al. 2016), and even in medical research and surveillance (Norton and Strauss 2013; Chunara et al. 2012; Collier et al. 2011).

Some user-reported data are generated from home, without a clinical interview or medical expert intervention. On one hand, these data record wide aspects of user life beyond hospitals and clinical tests, complementing health provider collected data and filling in the gap between clinical tests and treatments (PwC HRI Social Media Consumer Survey 2012). In some cases, user-reported data lead to new health definitions (Kallinikos and Tempini 2014). On the other hand, user-reported data have quality issues. Relevant literature reveals several reasons: lack of domain knowledge, privacy concerns, health status, and demographics. First, lack of medical knowledge can cause non-standard or incorrect terminologies and ambiguous or biased reference to side effects. Some users of PatientsLikeMe report symptoms not predefined in the system and provide unclear descriptions (Kallinikos and Tempini 2014). Second, online information privacy research has found that participants concerning their privacy may provide inaccurate or incomplete information as protective action (Son and Kim 2008). HVC users may be unwilling to share intimate feelings or personal concerns, especially on taboo topics (such as cancer and HIV) due to privacy concerns, which can lead to incomplete data, inaccurate data, or selectively reported data. Lastly, people of different age, health condition, and insurance type have different levels of willingness to share and engage in social media. The PwC HRI Social Media Consumer Survey (2012) found that senior citizens with poor health are least likely to share, trust, or engage in social media. As such, the collective information on HVC may be distorted. In addition to participation bias, health related emotional state could affect one's willingness to share his health information. The more negative one feels about his health, the more willing he is to give access to his health information (Anderson and Agarwal 2011). Health related emotions might lead to subjectiveness, inaccuracy, and incompleteness in reported data.

We look to the data quality and HVC design literature to address the data quality issues. Data quality can be investigated from four aspects: intrinsic, contextual, representational, and accessibility and security (Wang and Strong 1996). Promising platform design features include controls for anonymity (Kane and Ransbotham 2016; Leimeister et al. 2005); filtering and matching mechanisms for symptoms and other medical terms (Kallinikos and Tempini 2014); and openness about information usage and user benefits (Barrett et al. 2016).