WHAT CAN WE LEARN ABOUT THE MIDDLE EAST RESPIRATORY SYNDROME (MERS) OUTBREAK FROM TWEETS?

SUNMOO YOON*

School of Nursing, Columbia University Medical Center
New York, NY, 10032, USA

DA KUANG

Department of Mathematics, UCLA, Los Angeles, CA, 90095, USA

PETER BROADWELL

Digital Library, UCLA, Los Angeles, CA, 90095, USA

HAEYOUNG LEE

Department of Nursing, Hoseo University, Asan, South Korea

MICHELLE ODLUM

School of Nursing, Columbia University Medical Center
New York, NY, 10032, USA

ABSTRACT. Middle East Respiratory Syndrome (MERS, 메르스 in Korean) is an emerging deadly viral respiratory disease with no treatment. This study applied a triangulation approach of quantitative structure and content mining techniques while incorporating qualitative approaches guided by domain experts, to understand #MERS and #메르스 tweets. This study sought to gain insights about culturally-appropriate nursing activities for an emerging global acute disease management.

1. **Population health and Middle East respiratory syndrome (MERS).** Middle East respiratory syndrome (MERS, 메르스 in Korean) is a viral respiratory disease first identified in Saudi Arabia in 2012 with a 36% mortality rate. MERS has no vaccine or specific treatment. People in close contact with the contracted patients (e.g., care providers) or camels (e.g., travelers) are at high risk. Recently a MERS outbreak occurred in South Korea from May 26 to July 4 2015 resulting in 186 cases and 36 deaths. In public health, early detection and response is critical for emerging infectious disease outbreak as hospitals have established systematic infrastructures of rapid response teams. Social media has the potential to become a vehicle to assess public responses and to disseminate accurate community level information. This study characterized the cultural difference between #MERS and #메르스 tweets applying content and structure mining to gain insights of acute disease outbreaks.

2. **Methods.** A conceptual model of nursing and population health [3] guided our mining process. According to the model, upstream, population, health care system factors influence culturally appropriate MERS prevention activities and population level health outcomes. Tweets mentioning #MERS (English hash tag 543,836) and #메르스 (Korean hash tag 200,240) were randomly collected during the MERS outbreak from June 1st to July 4th, 2015 using NCapture. Sentiment analysis [2, 5] and topic modeling techniques including Latent Dirichlet allocation (LDA) [1] and Nonnegative Matrix Factorization (NMF) [4] were applied for content mining. Structure mining techniques were applied to detect macro (overall structure), meso (community detection) and micro (triad analysis) level communication networks [6] within Twitter. Temporal dynamic analysis were conducted in three phases: time 1 (T1: June 1-6), time 2 (T2: June 7-20), time 3 (T3:June 21-July 4, 2015).

3. **Results.**

3.1. **Content mining.** A total 93 different languages were detected in the #MERS (English hashtag) corpus. The average score of emotional valance was 5.3 on a 0 to 9 scale (0 being negative, 9 being positive). Arabic, Chinese, Russian tweets were noted as below average in frequency and French, German and Hindu tweets were noted as above the average in frequency. The #메르스 (Korean hashtag) corpus showed the emotional valance score 5.2 lower than the English hashtag #MERS corpus (Figure refpic1).

In #MERS tweet topic modeling, most tweets were related to population health outcome such as incidence, prevalence, and mortality reporting. Results applying LDA and NMF algorithms showed similar topics (e.g., fear, Samsung). #메르스 tweet content analysis using a triangulation approach of quantitative (NMF) and qualitative methods were illustrated in Figure 2; Green and blue colors represent health related topics, warm colors (pink) represent positive emotion, and grey colors represent negative emotion. Culturally sensitive ngrams (박근혜, 하트,보건소) were found in the #메르스 corpus.

3.2. **Structure mining.** Macro, meso and micro level network structure of tweets mentioning #MERS and #메르스illustrated in figure 3. Whereas the core part of each network in time 1, time 2, and time 3 decreased overtime in the English hashtag #MERS corpus, the size of the core part of each network in the Korean hashtag #메르스corpus increased over time. WHO was detected as a hub of Twitter communication in the early stage of MERS outbreak in the English hashtag tweets, a mayor of Seoul was detected as a hub in the Korean hashtag tweets. Korean Centers for Disease Control and Prevention (Korean CDC) appeared as a hub of the Twitter communication network in the later stage of the outbreak in the Korean hashtag corpus. Micro level network analysis showed that the English hashtag network mainly contained one-way communication type, the Korean hashtag network included a balanced communication type.

4. **Discussion.** Culturally-unique expressions were distinctly prominent in #메르스 tweets. On the contrary, most of #MERS tweets showed general case reports. Culturally-appropriate MERS prevention care activities in community health centers (systemin the conceptual framework) in the early stage of the outbreak might have been a crucial strategy for managing the deadly MERS virus in Korea. Twitter was largely used in South Korea as an effective information dissemination tool for 1) accurate disease prevention education and 2) emotional social support such
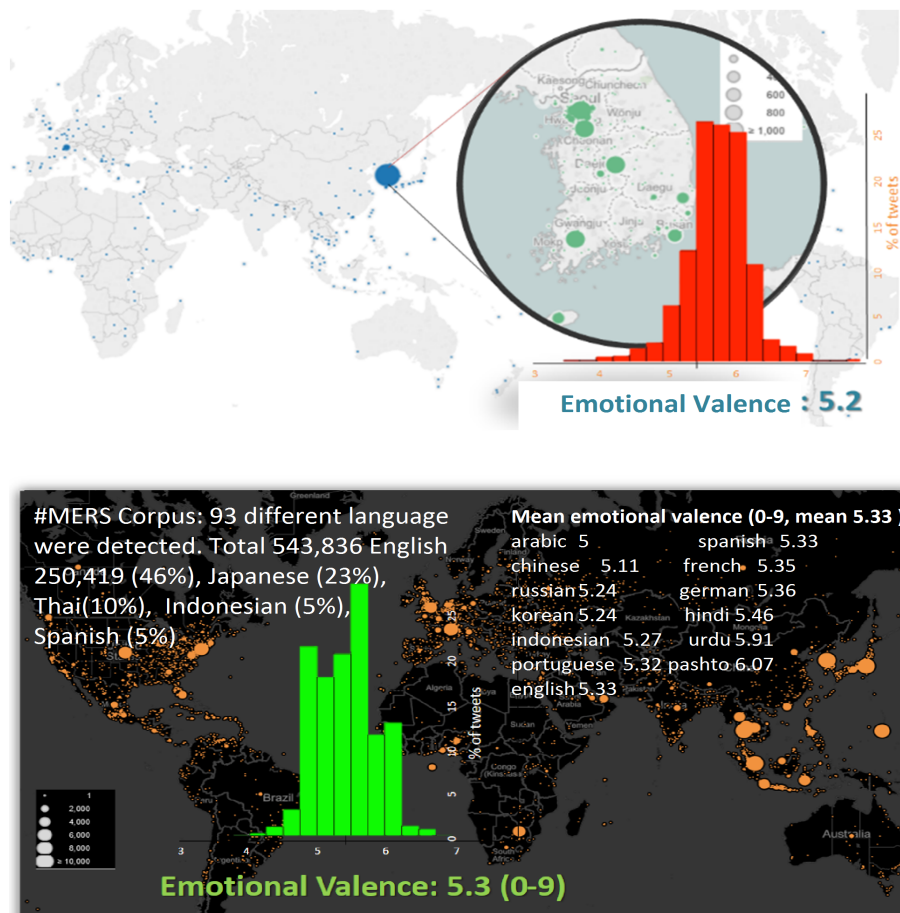
FIGURE 1. Sentiment of #MERS (top) #메르스 tweets (bottom)

as encouraging clinicians. In the sentiment analysis, emotional valance score of the Korean hashtag #메르스 tweets was slightly lower than the average. This unexpected results may be explained from our content and network analysis. Whereas negative emotions (e.g., fear, sad, and blaming the presidents) were prevalent in the early stage, positive tweets (e.g., encouraging clinicians and helping neighbors) were detected at the later stage. This implies that culturally sensitive stories in social media might have been helpful for emotional care during the devastating situation. Network analysis depicted that the hub of the communication was a Mayor of Seoul in the early stage when the negative emotions were prevalent. Meso level analysis revealed that the Korean CDC (Center for Disease Control and Prevention) appeared as a hub in later stages of the outbreak. Rapid response for fear management and education by the Korean CDC are recommended during the early stage of future globally emerging outbreaks. In terms of culturally appropriate disease prevention, the Twitter analysis showed that unscientific messages such as 'kimchi will cure MERS' were largely disseminated by politicians. Efficient strategies are needed to detect and correct culturally-sensitive misinformation. As meso and micro level analysis showed the established cohesive communities and ideal balanced
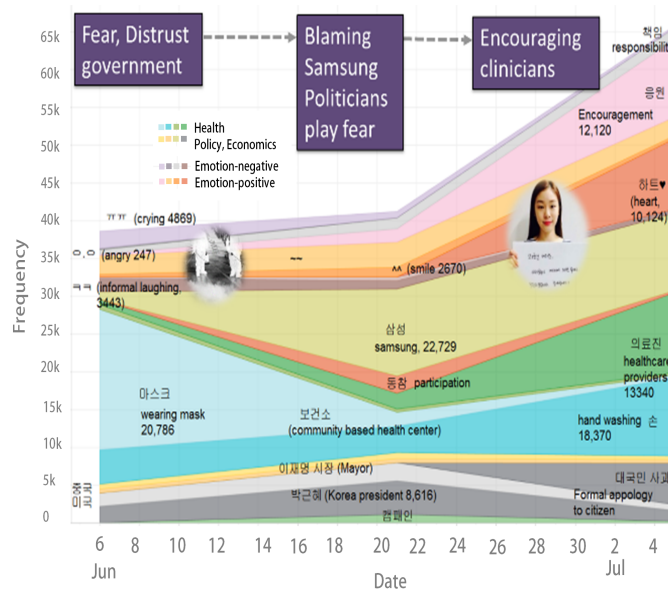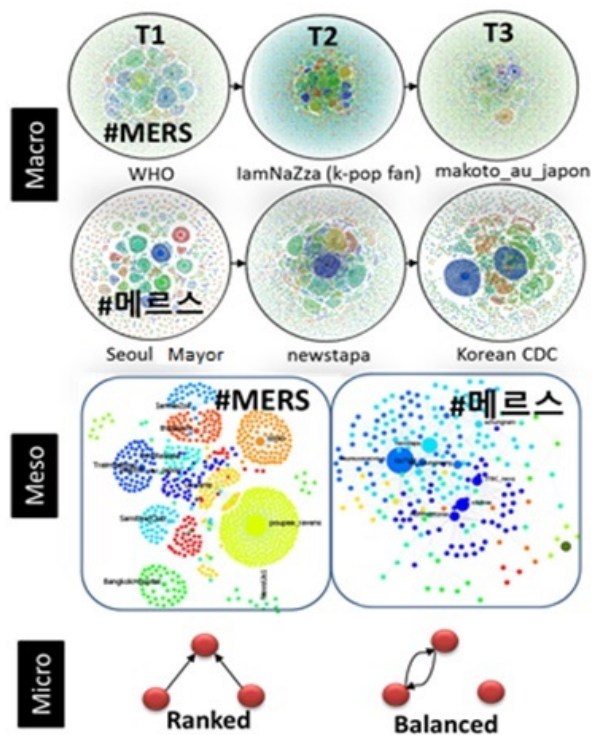
FIGURE 2. Content of #MERS and #메르스 tweets

FIGURE 3. Communication network of #MERS & #메르스 tweets

communication type as prevalent in the affected country, Twitter can be used as a complementary information dissemination vehicle which is universally accessible regardless of time or place.

5. **Conclusion.** This social media mining guided by Nursing and Population Heath framework was useful in gaining insights about culturally-appropriate nursing activities to decrease the burden of an emerging global acute disease. The triangulation of structure and content mining methods incorporating qualitative approaches guided by domain experts were helpful to comprehensively understand English and Korean MERS tweets.

REFERENCES

[1] D. M. Blei, A. Y. Ng and M. I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research*, **3** (2003), 993–1022.

[2] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss and C. M. Danforth, Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter, *PloS one*, **6** (2011), e26752.

[3] J. Fawcett and C. H. Ellenbecker, A proposed conceptual model of nursing and population health, *Nursing outlook*, **63** (2015), 288–298.

[4] D. Kuang and H. Park, Fast rank-2 nonnegative matrix factorization for hierarchical document clustering, in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2013, 739–747.

[5] M. Lui and T. Baldwin, Cross-domain feature selection for language identification, in *Proceedings of 5th International Joint Conference on Natural Language Processing*, Citeseer, 2011.

[6] S. Yoon and S. Bakken, Methods of knowledge discovery in tweets, in *NI 2012: Proceedings of the 11th International Congress on Nursing Informatics*, vol. 2012, American Medical Informatics Association, 2012.

E-mail address: Sunmoo Yoon: sy2102@columbia.edu
E-mail address: Da Kuang: dakuang@math.ucla.edu
E-mail address: Peter Broadwell: broadwell@library.ucla.edu
E-mail address: Haeyoung Lee: hylee@hoseo.edu
E-mail address: Michelle Odlum: mlo12@columbia.edu