Topical Mining of Malaria Using Social Media. A Text Mining Approach

James Boit  
*Dakota State University*

Omar El-Gayar  
*Dakota State University*

Follow this and additional works at: https://scholar.dsu.edu/research-symposium

Recommended Citation  
https://scholar.dsu.edu/research-symposium/14

This Book is brought to you for free and open access by the University Publications at Beadle Scholar. It has been accepted for inclusion in Annual Research Symposium by an authorized administrator of Beadle Scholar. For more information, please contact repository@dsu.edu.
Topical Mining of Malaria Using Social Media: A Text Mining Approach

Social Media Analytics

Abstract

Malaria is a life-threatening parasitic disease, common in subtropical and tropical climates caused by mosquitoes. Each year, several hundred thousand of people die from Malaria infections. However, with the rapid growth, and global reach of social media usage, a myriad of opportunities arises for extracting opinions and information regarding Malaria. Discourses on social media tend to focus on contemporary phenomena such as particular topical themes and discourses regarding Malaria extracted from the twitter corpus. Expected contribution and implications include shedding light and gaining insights on the public discussion surrounding Malaria inferred from multi-dimensional data in a manner that can provide directions for future research and inform public policy decisions.

Background

- Overview: According to the World Health Organization (WHO), over 400,000 people die of Malaria and a further estimated 219 million cases occur in 90 countries. As 92% of malaria cases and 93% of malaria deaths, Africa carries a disproportionate ratio of Malaria worldwide (WHO, 2019).

- Social Media: With over 3.5 billion people actively using social media, social media plays a significant role in creating novel ways for engaging in discourses, sharing narratives and opinions (Wearsocial, 2019).

- Challenge: Bureaucracy and lack of adequate resources continue to hinder effective detection and communication of infectious diseases in prone countries.

Literature Review

- Malaria is an infectious and lethal disease affecting most parts of the world with high prevalent incidences occurring in sub-Saharan Africa with an 81% of worldwide cases and 90% of reported deaths (Achrekar et al., 2010).

- Previous studies have focused on infectious diseases such as influenza, and Ebola cases (Chen et al., 2011). Other studies have focused on social media to detect and monitor influenza in Canada (Byrd et al., 2016), USA (Culotta et al., 2010), and UK (Lampous et al., 2010).

- Several studies in literature have investigated public concerns (Lazard et al., 2015), public discourses and discussions (Stefanidis et al., 2017; Orri et al., 2016), public health opinions (Karami et al., 2015), public discourses and discussions (Stefanidis et al., 2017; Orri et al., 2016), public health opinions (Karami et al., 2015), public discourses and discussions (Stefanidis et al., 2017; Orri et al., 2016), public health opinions (Karami et al., 2015), public discourses and discussions (Stefanidis et al., 2017; Orri et al., 2016), public health opinions (Karami et al., 2015), and trends (Kahatua et al., 2016; Alcino et al., 2015) in social media.

- While (Pak et al., 2010) illustrated that Twitter data can be an empirical source of information and vehicle to detect and track Influenza occurrences, (Prier et al., 2014) also demonstrated that Twitter can be a good source for discovering conversations and trending patterns in the health domain.

- Currently, non-traditional sources of data exist such as dedicated websites and social media platforms such as Facebook, and twitter where individuals readily share their ailments and illness experiences before deciding to seek medical services (Sharpe et al., 2016).

- Application of text mining in social media is an important technique in the extraction of timely and efficient information for use by public health practitioners and providers (Gomide et al., 2011).

Analysis and Discussion

- **Key words**: We created keywords to define the emergent categories.

- **Search Query**:

- **Results**:

- **Expected Contribution and Implications**: Contributes to enhancing understanding and communication of Malaria discourses on social media.

Conclusions

- **Summary**: Majority of discourses on Twitter converge around Drugs and symptoms with related conversations on travel, side-effects among others.

- **Limitations**: We used limited categories to analyze the initial emergent topical themes, perhaps further refinement and an expanded category may reveal much more breadth and depth covering Malaria discourses.

- **Expected Contribution**: Revealing the conversation patterns and trends shaping public opinion regarding Malaria discourses on social media.

- **Future Direction**: Motivate directions for future research, support large-scale geographical interventions and public policy decisions.

Methodology

- **Technology**: We used Crimson Hexagon Software, a powerful AI software to extract and analyze the topical themes relevant to the Malaria domain.

- **Corpus**: Twitter data with 1,168,437 Million relevant English-Language Tweets.

- **Technique**: Text mining models, specifically **Topic Modelling** was used for discovery of primary themes emerging from the corpus.

- **Development**: We developed and trained an **Opinion Monitor** to collect Tweets about Malaria between May, 2008– Dec, 2018.