Nurses and Midwives in the Digital Age
M. Honey et al. (Eds.)

doi:10.3233/SHTI210711

Web-Based Text Analysis of the Patient Safety Concerns of Various Healthcare Stakeholders

Insook CHO[a,1], Minyoung LEE[a] and Yeonjin KIM[b]
[a] *Department of Nursing, Inha University, Incheon, South Korea*
[b] *Department of Statistics, Inha University, Incheon, South Korea*

Abstract. Patient safety is a fundamental aspect of the quality of healthcare and there is a growing interest in improving safety among healthcare stakeholders in many countries. The Korean government recognized that patient safety is a threat to society following several serious adverse events, and so the Ministry of Health and Welfare of the Korean government set up the Patient Safety Act in January 2015. This study analyzed text data on patient safety collected from web-based, user-generated documents related to the legislation to see if they accurately represent the specific concerns of various healthcare stakeholders. We adopted the unsupervised natural language processing method of probabilistic topic modeling and also Latent Dirichlet Allocation. The results showed that text data are useful for inferring the latent concerns of healthcare consumers, providers, government bodies, and researchers as well as changes therein over time.

Keywords. Patient safety, natural language processing, topic modeling, healthcare stakeholders

1. Introduction

The recent rapid growth in user-generated text data shared in online communities, posting boards, and social media has made it possible to study and analyze language at an unprecedented scale. The analysis of user-generated texts on patient safety could reveal the topics that interest various stakeholders. However, the length of a specific text often exceeds the limit of what an individual person can read and process. Natural language processing (NLP) is a research and application area of computer studies that involves analyzing written or spoken language to help extract meaning from texts, and it has been applied in highly diverse disciplines and applications such as social media, political speeches, and physician discharge summarie [1]. Unsupervised NLP techniques, such as topic modeling, offer another method to understand free text without requiring the resources of supervised learning. Topic modeling, such as Latent Dirichlet Allocation (LDA), is a statistical approach to discover or identify topics associated with words or phrases [2].

In this study we applied the topic modeling technique to explore text data on patient safety collected from user-generated online texts. We were interested in how text data

[1] Corresponding Author, Insook Cho, PhD, RN, Department of Nursing, Inha University, Inharo 100 Namgu, Incheon 22212, Republic of Korea; E-mail: insook.cho@inha.ac.kr.

are distributed across the LDA topics, and in particular how this distribution can represent the specific concerns of various stakeholders.

2. Methods and Results

We divided the stakeholders into four groups: consumers (patients, caregivers, and families), providers (physicians, nurses, and professional healthcare organizations), governments (including legislative bodies and accrediting agencies), and researchers. We searched web-based communities and public sites for posting and news-sharing activities as well as opinions and information on patient safety. We found 18 representative sites for the stakeholder groups. We collected text documents that had timestamps indicating that they had generated on a posting board or in a newsroom or announcement between Jan. 2014 and Sep. 2018. Applying a series of data preprocessing steps using the KONLP package and manual review and filtering identified a document-term matrix of 2,487 documents and 2,933 words. We applied LDA in the topicmodels package of the R open-source software package [3], along with the ldatuning and log-likelihood statistics functions. The posterior distribution was estimated using Gibbs sampling and Markov-chain Monte Carlo simulations. We identified that the optimal number of topics was 41; Table 1 lists the top-3 topics according to stakeholder groups.

Table 1. Top three topics according to healthcare stakeholders inferred LDA modeling.

Stakeholder group	Top-three topics in each group
Consumer	Topic 3: Hospital infection control and newborn deaths at a neonatal intensive care unit
	Topic 28: MERS (Middle East respiratory syndrome) and visiting patients in hospitals
	Topic 30: Illegal surgeries by unlicensed persons
Provider	Topic 6: Institutional infection control
	Topic 11: Government policy
	Topic 32: Patient safety precaution alerts
Government	Topic 15: Enactment of legislation
	Topic 21: Systems for reducing damage due to drug side effects
	Topic 22: Infusion-related infections
Researcher	Topic 10: Institutional actions on patient safety
	Topic 14: Healthcare providers' perceptions of patient safety
	Topic 38: Communications about patient safety

3. Discussion and Conclusions

We found that infection control by medical facilities was the main common concern of healthcare stakeholders during a recent 5-year period in Korea. There were also trends in the changing concerns about illegal medical behaviors, legislation, and drug side effects between before and after the enactment of the Patient Safety Act in 2015. We further found discrepancies between the topics studied by researchers and the concerns of other stakeholders. The findings of and methods used in this study could form the basis of a bottom-up approach for national strategic planning about patient safety.

Acknowledgement

This work was supported by the Basic Research Program through the National Research Foundation of Korea (NRF-2016R1D1A1A09919502).

References

[1] Fong A, Ratwani R. An evaluation of patient safety event report categories using unsupervised topic modeling. Methods of information in medicine 2015;54(4):338–45.
[2] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine learning research 2003;3(Jan):993–1022.
[3] Grün B, Hornik K. Topicmodels: An R package for fitting topic models. Journal of statistical software 2011;40(13):1–30.