|  |
| --- |
| **Analysis of On-line News Articles on Women's Health Using Topic Modeling Technique** |
| **Background/Objectives**In this paper, we tried to understand the flow of interest in the field of women's health through analysis of news articles on healthcare. Researches on the analysis of news articles on women's health have been conducted in the past, but were limited to contemporary analysis and only one kind of disease. Therefore, we have collected and analyzed articles that have dealt with women’s health related topics in this paper. We would like chronologically to classify and examine how social issues related to women's health and diseases have changed by suddenly changing social environment.**Methods**The online news articles were collected from selected news web-sites and the saved data were preprocessed so that the analysis results can be accurately derived. Then the revised data by preprocessing were analyzed using LDA algorithm and major topics representing the news articles were extracted from LDA analysis results. Topic trends by year and period were analyzed using the extracted topics and inter-relationships of the extracted topics were analyzed too.**Results**From 1993 to 2000, 10 representative topics on women's health were identified. Topics of ‘Cerebrovascular disease’, ‘Skin health’, ‘Kidney disease’, ‘Dietary supplement’, ‘Thyroid disease’, and ‘Lifestyle disease & prevention’, were the newly emerging topics related to women's health since 2000. Looking at the main topics from 2006 to 2010, ‘Hair loss’, ‘Skin care’, and ‘Beauty treatment’ were new topics. Especially, in the topic related to ‘Lifestyle disease’, we found that several words related to prevention of infection against swine flu were extracted. Looking at the main topics from 2011 to 2015, ‘Infectious disease’ was a new topic. This also indicates that the social and personal interest in the MERS infection that caused the big wave for a while was high.**Discussion**The characteristics of the 1990s were that healthcare was the mainstream rather than specialized articles on topic areas. In the early 2000s, the topic related to beauty was comprehensive as a ‘Dietary supplement’. By the late 2000s, beauty related topics began to be subdivided into ‘Skin health’, ‘Hair loss’ and ‘Beauty treatment’. We could find that healthcare service was represented by key words for hospitals and support groups in the early 2000s. In addition, we could find the foundations of medical services expanded to society and government by 2010 because e-health has been developed.**Keywords**Women’s health, On-line news article, Topic modeling, LDA |