CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Social media has emerged as the most important platform for public opinion. Accompanied by the improvement of mobile internet, traditional mass media has lost much of its influence over the public. Under more relaxed audit conditions, social media reflects the public's genuine perspectives on various topics. It is essential to discern valuable information from posts that are presented of text and emojis. In the following sections, the deatils will explain how to filter posts by topics to identify related content. The project will annalyze and interpret genuine reactions of the posts. The analysis results will be summarized to draw conclusions..

2.2 Causes of Social Media Analysis

In the early days, social media analysis relied on computational linguistics and information retrieval systems, such as Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003). This model was primarily used with large text databases. However, as the variety of social media post types has increased and the limitations of older technologies became evident, contemporary researchers have turned to artificial intelligence (AI) to analyze social media content. The most commonly employed technologies now include machine learning techniques, such as deep learning, topic modeling algorithms, and natural language processing (NLP) methods, which are used to better understand the complexities of social media content.

2.3 Methods of Social Media Analysis

Latent Dirichlet Allocation (LDA) is a first-generation analysis model that gained popularity following its introduction in 2003. It was primarily applied to analyze early comments, tweets, and blogs in order to identify trends among public users (Blei & Lafferty, 2007).

Deep learning approaches represent advanced models based on deep learning techniques. The most commonly used techniques include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), both of which have been employed to analyze text and extract structures from social media posts (Yang et al., 2016).

Transformers and pretrained language models are based on transformer architecture, such as GPT (Generative Pre-trained Transformers). These models can be trained on specific datasets to derive more accurate topics from unstructured text and better capture context, even in less structured or more disorderly content.

2.4 Application of Social Media Analysis

One of the most popular applications today is the creation of topic models for social media to analyze public opinions on specific events, brands, or political issues. For example, a study by Karami et al. (2019) applied Latent Dirichlet Allocation (LDA) to Twitter data to assess public sentiment during the 2016 U.S. presidential election.

Another example is the use of topic modeling to analyze COVID-19-related discussions on Twitter, tracking public awareness, misinformation, and policy debates (Alamo et al., 2020).

2.5 Challenges and Future Research

As researchers seek to analyze social media posts, several challenges arise, such as low data quality. Posts are often written in a short, informal style, using internet slang, and incorporating emojis and hashtags, which complicates the pre-processing of data for topic modeling (Ruths & Pfeffer, 2014). Another significant issue is multimodal data, as social media posts include not only unstructured text but also images and videos. This presents a major challenge in extracting meaningful topics from such diverse forms of media (Xu et al., 2020).

Accompanied by the rapid development of AI technology, new models are now capable of handling more complex tasks in analyzing social media posts. For example, ChatGPT allows users to submit images, videos, and emojis to analyze topics within context. Enhancing the efficiency of these models will be a key focus of future research.

Reference

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. The 2007 IEEE International Conference on Data Mining, 17-24.

Karami, A., Baghery, A., & Shamsi, M. (2019). Public sentiment analysis on social media during the 2016 U.S. presidential election: A topic modeling approach. Social Media + Society, 5(2), 1-13.

Alamo, T., Arnaldo, J., & Sànchez, P. (2020). COVID-19: A large-scale social media analysis to understand the impact of the pandemic. Journal of Public Health, 42(4), 600-612.

Ruths, D., & Pfeffer, J. (2014). Social media for large-scale social research. Communications of the ACM, 57(11), 54-60.

Xu, K., Yadati, A., & Jansen, B. J. (2020). Multimodal social media analytics: Trends, techniques, and applications. Journal of the Association for Information Science and Technology, 71(10), 1231-1246.