BAB 1 Chapter 1 Introduction

1.1 Introduction

With the advancement of Internet technology and the popularity of smart mobile devices, social media has become an indispensable part of people's daily lives. The short text content such as tweets, microblogs, comments, replies, etc. posted by users on social media platforms not only reflects their immediate emotions and attitudes, but also has great research value. Whether it is for the business activities of enterprises (such as public relations) or the social governance of the government, the discovery and tracking of public opinion is very important. The core algorithm in public opinion analysis is the sentiment analysis algorithm.

The main creators of data on social networks are ordinary people. When they create content spontaneously, the content is diverse, including but not limited to short videos, short texts, pictures, emoticons, long texts or long videos. These are different from the traditional language environment. Therefore, this study will focus on the special language environment of social networks, collect and analyze data on online short texts on multiple social media in China, build a new word discovery system based on the language environment, track and collect data on multiple platforms for the same event at the same time, and use natural language processing technology and deep learning models for sentiment analysis.

1.2 Background of the Problem

With the popularity of social media, online public opinion plays an increasingly important role in shaping public opinion and influencing social events. For policymakers, social media is an important channel to understand public opinion and needs. Through deep learning analysis of large-scale text data from online

platforms such as social media and news comments, the public's reactions and emotional tendencies to specific policies or events can be quickly captured. This provides policymakers with real-time and comprehensive public opinion feedback, which helps to formulate policies that are more in line with public opinion. For companies, brand crises may break out at any time and have a serious impact on their reputation. By analyzing emotions on social media through deep learning technology, companies can promptly detect the spread of negative emotions and take corresponding measures for crisis management. This analysis can help companies understand consumer emotions, develop effective communication strategies, and mitigate the damage caused by crises.

In recent years, there has been an endless stream of research on online public opinion analysis. Li systematically shared the development and research status of Chinese public opinion research over the past few years in the "Current Status and Research Findings of Internet Public Opinion in China." in 2023. The core algorithm of public opinion analysis, sentiment analysis technology, has developed and iterated very quickly. From the earliest natural language models based on statistics, to deep learning models based on recurrent neural networks, to deep learning methods based on pre-trained models, and the research focus in the past two years, multimodal and multi-task sentiment analysis models. New models are increasingly focusing on sentiment analysis in complex language environments such as online texts, and the requirements for high-quality data sets are also constantly decreasing. For example, in 2023, Chen proposed a high-performance semi-supervised sentiment analysis method.

1.3 Statement of the Problem

At present, China's social media is mainly divided into three categories. The first is short text social media represented by Weibo, Tieba, and Zhihu. The main content is spontaneously created by platform users, including text and replies. The actual content includes short text and emoticons. The second category is vertical

social media represented by Xiaohongshu. The content is highly aggregated by field characteristics. The specific content is mainly pictures and texts, and comments and replies include short texts and emoticons. The third category is video websites represented by Bilibili, Kuaishou, and Douyin. The publishers are more formal and mainly create videos. Comments and replies are mainly short texts, including emoticons.

Existing social media research is limited to the first type of social media. There is no good research on the cross-domain issues of the second type of social media, and there is no special research on the third type. When the same event appears on different platforms at the same time, users on different platforms react differently. Traditional research has failed to effectively understand the differences and conduct unified analysis and processing. This study aims to build a cross-platform sentiment analysis algorithm, intending to achieve unified multi-platform data analysis from data collection and core algorithms.

1.4 Research Questions

(a) How to collect and organize data on the same event on different types of platforms with different structures?

(b) What are the differences in the content of network data from different types of media sources for the same event?

(c) Can a deep learning model be built to perform sentiment analysis on short online texts on multiple platforms at the same time?

1.5 Objectives of the Research

In response to the above problems, this study proposed the following research objectives:

(d) Build a program that can collect multi-platform data for a specific event.

(e) Analyze the data content of the same event on different platforms to find out the similarities and differences.

(f) Study the actual effect of different types of sentiment analysis models on the collected data? Is there room for improvement?

(g) Build a deep learning model that can perform efficient data analysis on short online texts on multiple platforms.

1.6 Scope of the Study

The research will use publicly available online datasets, including but not limited to SMP2020 and cnsent.

(h) The analysis will include at least one of the three types of social media.

(i) The analysis will include the impact of introducing traditional text sentiment analysis datasets on online short text analysis.

(j) The analysis will limit the content to Chinese social media.

(k) The analysis will include the impact of cross-language models, cross-language datasets, and sentiment analysis.

(l) The analysis will be limited to online short texts and will not include videos, pictures, or long texts.

1.7 Significance of the Research

This research can more effectively track and discover social public opinion, and solve the problem that the same event has different responses on different platforms, while traditional methods are ineffective. Ultimately, it can effectively track and discover the real impact of the same event on the Internet, which is beneficial to the government's social governance and corporate brand crisis management.

1.8 Structure of the Thesis

REFERENCES

Li, Xiaochen. (2023). "Current Status and Research Findings of Internet Public Opinion in China." Office Automation 28 (2): 16 – 18. https://d.wanfangdata.com.cn/periodical/bgzdh-bgsbygc202302005.

Haifeng Chen, Chujia Guo, Yan Li, Peng Zhang, Dongmei Jiang. (2023). "Semi-Supervised Multimodal Emotion Recognition with Class-Balanced Pseudo-Labeling." In MM '23: Proceedings of the 31st ACM International Conference on Multimedia, 9556 – 9560. https://doi.org/10.1145/3581783.3612864.