

A Design and Framework for Measuring Social Media Cyberbullying: Based on Text Sentiment Analysis

Zhou Fang^a

School of Media & Art Design, Guangzhou Huali College, Guangzhou 511325, China

Abstract. Cyberbullying is a focal issue of public concern. This study focuses on cyberbullying and innovatively introduces text sentiment analysis techniques into the research of specific social behaviors. Through this technique, a research framework for a cyberbullying scale in social media will be designed to explore the dissemination mechanisms and characteristics of cyberbullying in Chinese social media, and to develop a cyberbullying scale that is in line with the cultural background and characteristics of Chinese-speaking countries.

1. Introduction

Currently, the Internet is highly developed. We are accustomed to using the Internet and mobile social media to share various information and opinions. Although the Internet has brought convenience to human production and life, there are inevitable barriers in the process of human communication. Two-way or even multi-way communication is characterized by a dual coincidence. Therefore, inconsistencies in viewpoints easily give rise to problems such as cyberbullying and online harassment. Cyberbullying refers to malicious and aggressive acts against others using the Internet and electronic communication technology. This includes but is not limited to sending threatening, insulting, rumors, malicious comments, and other information through social media, text messages, emails, etc., which can cause victims to suffer psychological, emotional, or social harm in the virtual space. Currently, there is no unified recognition of the definition and scope of cyberbullying. People have different views on cyberbullying, and there is much controversy over its definition and classification. However, there is a generally recognized consensus on cyberbullying: it is a negative behavior that causes psychological and physical harm and to some extent affects the lives and work of the victims[2].

Social media platforms are one of the main venues for cyberbullying, with a massive amount of user data. Any internet user can easily harm or bully others by posting malicious comments or spreading rumors in online environments such as social media. Compared to traditional face-to-face bullying, cyberbullying provides perpetrators with a quicker and more convenient way to carry out bullying behaviors. Additionally, due to the anonymity provided by the internet, cyberbullying actions are more covert, and every social media user may face the threat of cyberbullying at any time. [5]Cyberbullying often leads to emotional responses such as anger,

depression, and anxiety, and can even result in serious consequences like social difficulties and suicide. According to a recent population-based study, the suicide rate (34.4% vs. 21.6%) and self-harm percentage (32.8% vs. 22.3%) of cyberbullying victims are higher than those of traditional bullying victims[1,10].

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in Table 2 should be used.

2. The Current Study

In order to study and assess the extent of cyberbullying, researchers have developed various scales to measure individuals' perceptions, experiences, and impacts when facing cyberbullying. These scales can help scholars and practitioners understand the scale, types, and impact of cyberbullying, and provide a basis for developing effective prevention and intervention strategies.

2.1. Cyberbullying and Scale

Although cyberbullying is very common, there is currently a lack of universally accepted definitions and assessment tools for cyberbullying, as well as limited support in terms of prevention, treatment, and intervention measures. Currently, there are two main approaches to measuring cyberbullying based on the existing CBB measurement tools. The first is through questionnaires, such as the Cyberbullying Behavior Scales developed by Willard (2007) and Yang (2012)[6]. The second approach involves using machine learning algorithms and natural language processing techniques to measure, identify, and assess cyberbullying behaviors by mining data[7].

First of all, in the measurement of cyberbullying, the survey method (cyberbullying questionnaire) is still the main method used by most domestic and foreign

^a neo6199@126.com

researchers. Most researchers write questionnaires based on existing scales. Currently, the widely used scales include QoCB (Arıcak et al., 2008), CBQ (Calvete et al., 2010), and CBI (Erdur-Baker & Kavrut, 2007)[6]. However, based on the current research on scales, there are many issues, such as the widespread use of single-dimensional scales, undetermined factor structures, and the use of open-ended inquiries to obtain information, lacking reliability in the information.

Secondly, based on the measurement of machine learning algorithms, most researchers in the field of computer science study the detection of cyberbullying based on technological aspects. There are several problems in identifying cyberbullying behavior based on current machine learning algorithm measurements:

(1) Cyberbullying lacks a universally accepted definition, which varies across different cultures, societies, and age groups.

(2) There is a lack of good and authentic dataset during the detection process, which also needs to address cyberbullying samples defined by relevant ethics and laws.

(3) There is no consistent and reproducible evaluation method.

(4) Suitable algorithms need to be used to deal with the complex and diverse nature of cyberbullying[8].

2.2. Cyberbullying and Text Sentiment Analysis

The research design of this study will use text sentiment analysis as a research method and technique in natural language processing to capture and analyze cyberbullying language texts in social media, in order to identify and classify the negative emotions and aggressive language contained in cyberbullying behaviors.

Text sentiment analysis, also known as sentiment orientation analysis or opinion mining, is the process of extracting information from user opinions. It involves analyzing text, audio, and images to understand people's viewpoints, opinions, attitudes, and emotions. It is an important component of natural language processing technology in machine learning.

Based on the detection method of emotions and sentiments, this method emphasizes the types of emotions and sentiments conveyed by the text, including positive, negative, neutral, and considers the content of the text in order to accurately identify offensive content on social media[3]. Although strong emotions expressed through text are important indicators, they are still rarely sufficient to accurately identify whether the content is unpleasant, as they may fail to recognize sarcasm.

In existing research, the Sentiment-Based Cyberbullying Detection (SCID) framework proposed by Dani et al., as well as the method of social network analysis for identifying cyberbullying proposed by Ting et al., are able to some extent to recognize cyberbullying behavior and detect its impacts by examining emotional information displayed in social media. Additionally, Watanabe et al. have even developed a machine learning algorithm with an accuracy of 87.4% for binary classification (aggressive vs. non-aggressive) and 78.4% for ternary classification (hate, aggression, and

neutrality)[4]. However, these frameworks and methods have not explored the influence of hidden ironic information on the detection of cyberbullying behaviors.

3. Research Methods and Design

Based on the current research status, there is relatively more study on cyberbullying in Western countries, while there is relatively less research on cyberbullying in Asian countries. A recent cross-cultural study measured the attitudes towards cyberbullying in seven countries including Australia, Brazil, China, Germany, Japan, Singapore, and the United States. The results showed that all countries had a significantly low positive attitude towards cyberbullying, with China scoring the lowest. Therefore, this study will focus on cyberbullying in China and explore the phenomenon of cyberbullying in different cultural backgrounds. This aims to contribute to internet governance and provide insights on how individuals can cope with the threats and challenges of cyberbullying[9].

3.1. Research objectives

This research study has the following three objectives:

1. Investigate the mechanisms of cyberbullying propagation on Chinese social media: Collect social text data to analyze the characteristics, patterns, and mechanisms of cyberbullying behavior on social media. Use text sentiment analysis techniques to explore the language and emotional expressions of cyberbullies, in order to understand the psychological factors and motivations behind cyberbullying.

2. Develop a scale for measuring cyberbullying on social media: Based on existing research and empirical analysis, design and develop a scale specifically for measuring cyberbullying on social media. This scale will cover various types of cyberbullying behaviors, such as verbal insults, malicious comments, and privacy exposure. Additionally, employ sentiment analysis techniques to assess the intensity of negative emotions in language, thereby enhancing the accuracy and comprehensiveness of measuring cyberbullying behaviors.

3. Explore the relationship between cyberbullying and mental health: Conduct a survey using the developed scale and analyze the collected data to study the impact of cyberbullying on the mental health of victims. By comparing psychological indicators, such as depression and anxiety, between cyberbullying victims and non-victims, examine the psychological stress and negative emotions caused by cyberbullying. Additionally, propose corresponding coping strategies and intervention measures.

3.2. Research Design & Framework

This study includes data collection, data processing and data analysis, scale design, scale testing and sampling, scale refinement, and scale validation.

3.2.1 Data collection

Use web crawlers to retrieve text content related to cyberbullying from social media platforms such as Weibo. This process includes selecting data sources and collecting data. For the selection of data sources, it is important to choose representative social media comments, messages, or bullet screen texts that contain a high amount of cyberbullying content. Following the methods proposed by McHugh et al., the data collection should focus on posts related to cyberbullying over a time span of at least one month.

3.2.2 Data processing and analysis

After data collection, researchers obtain a certain amount of content related to cyberbullying, which includes rich textual information. In order to analyze these texts effectively, researchers need to perform data processing, including the following aspects:

(1) Data cleaning: Since the collected content consists mainly of Chinese vocabulary, researchers need to use the Python programming language and the Chinese word segmentation tool, jieba, to effectively identify cyberbullying vocabulary.

(2) Construction of datasets: The most important backup in this step is to construct datasets that meet the requirements for cyberbullying detection.

(3) Sentiment analysis: This study requires the use of sentiment analysis models based on natural language processing techniques to label words in the text into psycholinguistic categories, including identifying positive and negative emotions, as well as anxiety, anger, and sadness, especially words associated with aggression and hatred.

(4) Content analysis: This study can utilize natural language processing models to analyze cyberbullying words, such as determining the frequency of their occurrence and calculating the relevance and importance of the content to cyberbullying.

In this study, data analysis includes two aspects: sentiment analysis and content analysis.

(1) Results of sentiment analysis: In sentiment analysis, researchers need to label words in the text into psycholinguistic categories and then conduct statistical analysis on the sentiment categories.

(2) Results of content analysis: In content analysis, researchers need to use natural language processing models to analyze cyberbullying words, calculate the relevance and importance of the content to cyberbullying, and address other related issues.

3.2.3 Scale Design

The scale design of this study first identifies the individuals who may experience cyberbullying among social media users as the subjects. Secondly, a series of items describing cyberbullying behaviors need to be designed based on the selected theoretical framework, covering different types of cyberbullying behaviors. Finally, specific rating criteria need to be assigned to each

item, and the presentation format of the scale needs to be determined.

3.2.4 Scale Testing and Snowball Sampling

After determining the scale items, scale testing needs to be conducted. First, an initial sample is obtained through snowball sampling to ensure representative and diverse samples. The selection of samples needs to consider the demographic characteristics of the target population, especially age and occupation categories. The sample size should not be less than 280 to ensure the validity of the testing. Secondly, the designed scale is used in the sample for participants to rate based on their own experiences. Thirdly, the rating data of the participants need to be collected, along with their feedback and suggestions. Fourthly, preliminary analysis of the testing data needs to be conducted using statistical methods such as descriptive statistics and factor analysis to prepare for the subsequent evaluation of the reliability and validity of the scale[11].

3.2.5 Scale Refinement

Revised and deleted items in the scale based on the pilot test results and participants' feedback to ensure accuracy and reliability. Next, sorted and grouped the items according to different types of cyberbullying based on the revised scale for further analysis and interpretation. Lastly, considering whether the scale needs to be retested with a new sample to achieve scientific validity.

3.2.6 Scale Validation

Reliability and validity of the scale are important indicators of the reliability and effectiveness of quantitative research data. For scale reliability, internal consistency statistical methods (such as Cronbach's alpha coefficient) can be used to examine the reliability of the scale and assess the consistency among its items. For scale validity, methods such as correlation analysis and comparative group analysis can be used to evaluate the correlation and differences between the scale and other relevant variables, thus validating the scale.

3.3. Design & Dimensions of Cyberbullying Scale

Scale design and dimensions of cyberbullying:

The scale for measuring cyberbullying behaviors can be designed according to multiple dimensions. Based on the characteristics of social media, the dimensions and potential questions for the scale are presented in Table 1.

Table 1. Dimensions & Questions of Cyberbullying Scale.

Dimensions	Questions
Violent Tendency Assessment	Q1:Have you ever posted offensive or abusive comments targeting specific individuals or groups on social media? Q2:Have you ever intentionally spread false information or inappropriate content through social media to infringe on the rights of others or cause distress?

Social Adaptation Disorder Assessment	Q3:Have you ever engaged in arguments or even malicious attacks with someone on social media because of their comments? Q4:Have you frequently experienced being blocked, having your posts deleted, or being tagged as a malicious account by other users on social media?
Mental Health Assessment	Q5:Have you ever been severely mentally harmed due to defamation or malicious attacks on social media? Q6:Have you experienced anxiety, depression, or other negative emotions due to cyberbullying?
Behavior Control Assessment	Q7:Do you often deliberately choose or avoid certain topics when posting on social media to avoid being attacked by other users? Q8:Are you able to control yourself and avoid frequently making aggressive or violent remarks on social media?
Moral Judgement Assessment	Are you aware that your behavior on social media can harm the interests of others? Do you consider the feelings of other users when controlling your words and actions on social media?

4. Conclusion

This article aims to develop a scale for measuring cyberbullying behavior on social media networks based on text sentiment analysis techniques and explore it in order to better understand and address the phenomenon of cyberbullying. In this regard, based on the current research on cyberbullying, this article will focus on how to construct the research objectives and research priorities of this article based on the text sentiment analysis model, thereby exploring the research framework for developing a cyberbullying scale based on sentiment analysis, and designing the research accordingly. In addition, this study still has some issues. Firstly, it needs to address the definition and qualitative issues of cyberbullying, but there is currently no basic consensus in the research. Secondly, this study needs to consider the national conditions and environment of Chinese-speaking countries like China, as most of the research on measuring cyberbullying is concentrated in the United States and European countries. Therefore, the previous tools may not take into account the characteristics and environment of Asian countries, which is also related to the differences in language use. These issues need to be addressed in future practical research.

Acknowledgements

This research was supported by Youth Research Project of Guangzhou Huali College, grant number:HLKY-2022-SK-06

References

- Lu, S., Zhao, L., Lai, L., Shi, C., & Jiang, W. (2022). How Do Chinese People View Cyberbullying? A Text Analysis Based on Social Media. *International Journal of Environmental Research and Public Health*, 19(3), 1822.
- Vismara, M., Girone, N., Conti, D., Nicolini, G., & Dell’Osso, B. (2022). The current status of Cyberbullying research: a short review of the literature. *Current Opinion in Behavioral Sciences*, 46, 101152.
- Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P. C., Carvalho, J. P., Oliveira, S., Coheur, L., Paulino, P., Veiga Simão, A. M., & Trancoso, I. (2019). Automatic cyberbullying detection: A systematic review. *Computers in Human Behavior*, 93, 333–345. <https://doi.org/10.1016/j.chb.2018.12.021>
- Atoum, J. O. (2020). Cyberbullying Detection Through Sentiment Analysis. 2020 International Conference on Computational Science and Computational Intelligence (CSCI). <https://doi.org/10.1109/csci51800.2020.00056>
- Guo, S., Liu, J., & Wang, J. (2021). Cyberbullying Roles Among Adolescents: A Social-Ecological Theory Perspective. *Journal of School Violence*, 1–15. <https://doi.org/10.1080/15388220.2020.1862674>
- Zhang,Z.Z. (2021) . The concept, dimensions, and scale development of game bullying. *International News*, 43 (10), 69-97.
- Thomas, H. J., Scott, J. G., Coates, J. M., & Connor, J. P. (2018). Development and validation of the Bullying and Cyberbullying Scale for Adolescents: A multi-dimensional measurement model. *British Journal of Educational Psychology*, 89(1), 75–94. <https://doi.org/10.1111/bjep.12223>
- Lozano-Blasco, R., Quilez-Robres, A., Rodriguez-Araya, R., & Casanovas-López, R. (2022). Addiction to New Technologies and Cyberbullying in the Costa Rican Context. *Education Sciences*, 12(12), 876.
- Lee, E. B. (2016). Cyberbullying. *Journal of Black Studies*, 48(1), 57–73.
- Escortell, R., Aparisi, D., Martínez-Monteagudo, M. C., & Delgado, B. (2020). Personality Traits and Aggression as Explanatory Variables of Cyberbullying in Spanish Preadolescents. *International Journal of Environmental Research and Public Health*, 17(16), 5705. <https://doi.org/10.3390/ijerph17165705>
- Patchin, J. W., & Hinduja, S. (2015). Measuring cyberbullying: Implications for research. *Aggression and Violent Behavior*, 23, 69–74. <https://doi.org/10.1016/j.avb.2015.05.013>