

Technical Research on Social Media Text Stance Detection

Jiayi Zhao

Harbin Engineering University, Harbin, China

Email: zhaojiayi11213@hrbeu.edu.cn

Manuscript received April 28, 2026; accepted May 28, 2026; published June 15, 2026

Abstract—With the widespread adoption of social media, the speed and reach of information dissemination have expanded dramatically in unpredictable ways. This is not merely a simple phenomenon; it is fundamentally changing the way we acquire and process information. Stance detection in social media text is therefore particularly important, especially in rapidly changing and emotionally charged online discussions. Stance detection goes beyond mere sentiment analysis; it involves profoundly analyzing the complex attitudes, stances, and implicit biases embedded in the text. In recent years, with the rapid development of Large Language Models (LLM) and deep learning technologies, traditional stance detection methods have gradually been replaced by more complex and sophisticated techniques, particularly in large-scale text processing, where deep learning plays an increasingly prominent role. This paper will explore the construction of stance data, the evolution of model paradigms, and the issue of generalization in real-world applications. The challenge of stance detection lies not only in accurately identifying stances but also in integrating various factors—such as emotion, culture, and social context—into the model. This paper will analyze this from multiple levels, including data construction and technological changes, and propose new theoretical paths to improve the practicality and credibility of this field.

Keywords—stance detection, social media text, large language models, deep learning

I. INTRODUCTION

The core task of stance detection is to identify implicit stances or sentiments in a text, a process that is more complex than simple sentiment classification. Social media texts are highly subjective and context-dependent; therefore, stance detection is not merely a matter of sentiment, but also requires understanding the underlying social background and changing contexts [1, 2]. Users on social media often express their attitudes through language with strong biases, reflecting the collective cognition and emotions of society on a specific topic. The complexity of this process lies in the fact that texts on social platforms must not only cope with linguistic diversity but also account for the subtle differences between different cultures, groups, and events.

One of the challenges of stance detection is extracting accurate sentiment information from complex and diverse social media data. Today, researchers no longer rely solely on simple keyword-based methods or sentiment dictionaries but instead use deep learning and large language models to delve deeper into the emotional connotations of texts [3, 4]. The research scope of stance detection covers multiple aspects, from data collection to model optimization and evaluation methods. This article will explore the evolution of these aspects, starting with the complexity of position data and the diversity of construction methods, analyzing the application of traditional and modern technological paradigms, and finally discussing how to solve the problems of generalization and evaluation in practical applications.

II. DATA FOUNDATION THEORY: THE ROOTS OF STANCE REPRESENTATION AND THE THEORETICAL EVOLUTION OF DATA CONSTRUCTION

A. The Inherent Complexity of Position Data: Subjectivity, Context Dependence, and Dynamism

Stance data in social media texts are not only highly subjective but also deeply influenced by context and cultural background. The expression of stance is often closely related to specific topics, backgrounds, and social environments; therefore, effectively capturing these emotional changes in different contexts has become a major challenge in stance detection [5]. Traditional sentiment analysis methods often struggle to cope with this complexity and are unable to accurately capture the subtle differences underlying stances.

In recent years, many studies have begun to utilize contextual information and features to enhance stance recognition capabilities. This approach can, to some extent, compensate for the limitations of sentiment analysis, thereby making stance detection more accurate and efficient.

B. Theoretical Evolution of Data Construction Methods

1) Early Paradigm (Around 2016)

Before 2016, stance detection relied primarily on feature engineering and shallow machine learning methods. While these methods relied on manually designed features, such as vocabulary and syntax, they often fell short when faced with rapidly changing social media text [6].

2) Development Paradigm

With the development of deep learning technology, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) became new research directions. The introduction of deep learning models solved the limitations of traditional methods in processing complex text, significantly improving the accuracy of stance detection. However, with the application of these methods, researchers also began to face challenges in data annotation and model generalization [7].

Driven by Large Language Models (LLMs) such as GPT-3, stance detection research entered a new stage. Through large-scale pre-training and fine-tuning, LLMs can handle more complex and diverse social media texts and extract more nuanced stance-related information from them. These models can not only identify positions in text, but also generate explanatory content related to positions, further improving the transparency and interpretability of position detection [8, 9].

III. MODEL PARADIGM THEORY: THE TECHNOLOGICAL PHILOSOPHICAL SHIFT FROM REPRESENTATION LEARNING TO INFERENCE MODELING

A. The Stage Based on Feature Engineering and Shallow Learning

In this stage, stance detection models mainly relied on manually designed features and shallow machine learning algorithms, such as Support Vector Machines (SVMs) and random forests. These methods typically determine stance by identifying some keywords or structured features in the text. Although these methods performed reasonably well in specific tasks, they showed significant limitations for complex social media texts [10]. With the increasing diversity and openness of social media data, traditional feature engineering methods began to fall short, failing to effectively capture deep-seated sentiment and stance information in the text.

B. Deep Representation Learning Paradigm

With the widespread application of deep learning, stance detection entered a new stage of development. Models based on technologies such as word embeddings and Long Short-Term Memory (LSTM) networks became mainstream. At this time, the design of the model no longer relied on manually selected features, but rather inferred stance by automatically learning representations in the text. This shift significantly improved the accuracy of stance detection. However, despite the excellent performance of deep learning methods, challenges remain in processing long texts and understanding complex contexts [5].

C. A New Paradigm Driven by Large Language Models (LLM)

In recent years, the rise of large language models has brought new opportunities for stance detection. Through large-scale pre-training and fine-tuning, large language models can not only effectively process complex social media texts but also flexibly adjust stance predictions according to contextual changes. The advantage of this paradigm is that the model can handle more diverse text inputs while not relying on cumbersome feature engineering [8, 9]. In addition, LLMs have high transparency and interpretability, providing new possibilities for the further development of stance detection.

IV. TASK GENERALIZATION THEORY: EXPLORING GENERALIZATION THEORY TOWARDS REAL-WORLD APPLICATIONS

In this chapter, we will delve into task generalization, focusing specifically on the challenges and opportunities in real-world applications. First, we will analyze the core problems facing task generalization, revealing how to ensure the model maintains high performance across tasks and domains. Next, this chapter will examine the current theoretical foundations of task generalization, explore existing theoretical frameworks, and propose some possible theoretical breakthroughs aimed at facilitating the transfer of models from laboratory research to practical applications.

A. The Essence of Generalization Challenges

This section will focus on the essential challenges facing task generalization, particularly how to maintain the stability and efficiency of models across different types of data or tasks. We will explore how to effectively address the generalization problem in multi-task learning, cross-domain transfer, and low-resource environments. By analyzing existing challenges and difficulties, this section will lay the theoretical foundation for subsequent solutions.

The generalization problem of position detection tasks, in short, is how to maintain model performance across different types of data or tasks. Although deep learning and large language models have made significant progress on social media texts in some specific domains, their stability is often difficult to guarantee when applied to multi-task or cross-domain applications. The core of this problem lies in the vast differences in the types and expressions of social media texts, especially on cross-cultural, cross-linguistic, or cross-domain social platforms, where the forms of stance representation vary greatly [6, 7].

Furthermore, the diversity of social media platforms, including differences in language, region, and cultural background, further increases the difficulty of generalizing stance detection. To address these challenges, researchers have proposed several new methods, such as cross-domain transfer learning and data augmentation techniques, which can effectively improve the generalization capability of models, especially in low-resource language and cross-domain settings [10].

B. Critique of Theoretical Paths to Improve Generalization

This section reviews the main theoretical paths for improving task generalization and examines their advantages and disadvantages from multiple perspectives. It focuses on how to address the generalization bottleneck through few-shot learning, transfer learning, and other methods, as well as the potential and limitations of these methods in practical applications. Furthermore, this paper explores innovative ideas at the theoretical level to seek new paths to further improve model generalization. Improving the generalization ability of stance detection models can be approached from several aspects: First, few-shot learning and semi-supervised learning provide new ideas for solving the problem of scarce data labels. These methods can train effective models on a small amount of labeled data, thereby improving the model's performance on unknown data [11]. Second, cross-domain transfer learning methods have gradually become important means of improving generalization capability. By borrowing knowledge from different domains, transfer learning can effectively improve the model's performance on specific tasks, but its universality in multi-task and cross-domain applications still needs further verification [9].

V. RESULT CREDIBILITY THEORY: THEORETICAL FRONTIERS OF INTERPRETABILITY, EVALUATION, AND ETHICS

This chapter focuses on the credibility of stance detection models, specifically on interpretability, evaluation criteria, and model ethics. With the widespread adoption of deep

learning models in practical applications, traditional evaluation methods can no longer fully reflect the model's true performance in complex situations. Therefore, this chapter will explore how to improve evaluation theory from multiple dimensions and examine the profound impact of interpretability and ethical issues. This theoretical framework aims to provide a more comprehensive and robust reference for the future development of stance detection.

A. Beyond Accuracy: Evaluation Theory for Stance Detection

This section examines the limitations of traditional stance detection evaluation criteria, particularly metrics such as accuracy, precision, and recall. We will explore why these criteria fail to adequately reflect the model's true performance in complex social media text environments and introduce some new evaluation methods proposed in recent years, such as interpretability, model transparency, and bias analysis. These methods will help researchers better understand model behavior and provide theoretical guidance for future optimization. Furthermore, this section will delve into how to balance model accuracy and fairness, emphasizing the importance of avoiding bias in practical applications.

Traditional stance detection evaluation criteria typically focus on metrics such as accuracy, precision, and recall. However, these criteria do not adequately consider the differences in model performance in complex situations. Especially when dealing with social media text, traditional evaluation methods may not accurately reflect the true capabilities of the model. Therefore, in recent years, several more refined evaluation criteria have been proposed, such as the interpretability of stance identification, model transparency, and bias analysis [5, 6]. These methods help researchers better understand the behavior of the model and provide valuable guidance for further model optimization.

Furthermore, with the widespread application of stance detection technology, the ethical issues of the model have gradually attracted the attention of the academic community. How to avoid bias in the process of stance identification, especially discrimination against specific groups or viewpoints, has become an important part of current evaluation criteria [8]. Therefore, the development of stance detection models should focus not only on their accuracy but also on their fairness, transparency, and social responsibility.

B. Evolution of Interpretability Theory

With the application of deep learning and large language models, interpretability research has become a research hotspot in the field of stance detection. This section analyzes the evolution of interpretability issues, from traditional "black box" models to the interpretable artificial intelligence (XAI) technologies introduced in recent years, and explores how attention mechanisms, visualization tools, and other methods can help reveal the model's decision-making process and internal mechanisms. By analyzing the advantages and challenges of these emerging methods, this paper explores approaches to improving the transparency and credibility of stance detection models, thereby laying a theoretical foundation for future model optimization.

With the application of deep learning and large language

models, research on interpretability has gradually become a hot topic in stance detection research. In traditional "black box" models, the reasoning process usually lacks transparency, making it difficult for users and developers to understand the model's decision-making basis [7]. These methods help reveal the model's decision-making process and internal mechanisms, thus supporting the further optimization of stance detection models [10]. These methods can help reveal the model's decision-making process and internal mechanisms, providing support for the further improvement of stance detection models [10].

VI. CONCLUSION

While existing research paradigms have achieved certain progress in stance detection, several limitations still exist. First, most studies rely on supervised learning, which requires a large amount of labeled data, potentially unavailable in certain domains or tasks. Second, existing models primarily focus on social media texts in English and other mainstream languages, with relatively few studies on stance detection in low-resource languages. Third, although large language models perform well across various tasks, they typically demand substantial computational resources, and their deployment in specific domains remains challenging [7].

Future research directions may focus on the following aspects: First, cross-domain transfer learning and few-shot learning will play a crucial role in improving the generalization of stance detection. By leveraging data from different domains or small amounts of labeled data, researchers can develop more efficient and universal stance detection models. Second, with the rise of multimodal data, future stance detection may extend beyond text to include comprehensive analysis of multiple information sources such as images and audio. Furthermore, ethics and fairness in stance detection models will remain key research issues, particularly in practical deployment on social media platforms [10].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Y. S. Sun, "Research on stance detection methods for public opinion in personal social media," Beijing University of Posts and Telecommunications, 2025.
- [2] M. Z. Guo, "Research on large language model-driven stance detection methods for social media texts," University of Science and Technology of China, 2025.
- [3] J. H. Li, L. Shi, M. Ding *et al.*, "Stance detection of social media texts based on large language models," *Journal of Frontiers of Computer Science and Technology*, vol. 19, no. 05, pp. 1302–1312, 2025.
- [4] J. W. Yuan, "Research on low-resource text stance detection methods for social media," Northeast Forestry University, 2024.
- [5] Y. J. Fu, "Research on stance detection and its interpretability in social media," Shanxi University, 2023.
- [6] J. H. Yuan, "Research on text stance detection technologies in social media," Harbin Institute of Technology, 2023.
- [7] Y. Q. Deng, "Research on stance detection for social media," Guilin University of Electronic Technology, 2023.
- [8] Y. L. Geng, "Research on stance detection for social media texts," Dalian University of Technology, 2023.
- [9] B. Liang, "Research on aspect-level sentiment analysis and stance detection for social media," Harbin Institute of Technology, 2022.

- [10] Z. Z. Su, Y. Y. Xi, Y. F. Chen *et al.*, "Data augmentation method for social media stance detection," *Journal of Information Engineering University*, vol. 23, no. 01, pp. 58–65, 2022.
- [11] Y. H. Fu, "Research on few-shot stance detection methods for social media texts," Harbin Institute of Technology, 2021.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](#)).