The interactive development of computational models and multimodal discourse analysis theory

Qunli Xie

Software Engineering Institute of Guangzhou

qunlixie98@163.com

Abstract. With the rapid advancement of computational technologies and the in-depth exploration of multimodal discourse analysis theory, the intersection of these fields has become a frontier area in both technology and linguistics research. This article examines the application of computational models in multimodal discourse analysis and analyzes how innovations in multimodal discourse analysis theory drive advancements in computational models. Through empirical studies and theoretical discussions, the paper reveals the interaction mechanisms between computational models and multimodal discourse analysis in practice and highlights their theoretical complementarity, providing new perspectives and methodologies for future research. The findings show that effective computational models enhance the accuracy and depth of multimodal analysis, while theoretical innovations in multimodal discourse analysis propel computational models toward greater efficiency and adaptability.

Keywords: Computational Models, Multimodal Discourse Analysis, Interactive Development, Linguistics

1. Introduction

In the information era, the integration of diverse media forms such as text, images, sound, and video has become increasingly prevalent. This trend calls for a re-evaluation of traditional discourse analysis methods. Multimodal discourse analysis, as an emerging analytical framework, allows researchers to interpret information comprehensively through visual, auditory, and textual signals. The continuous refinement of its theories and methods is expanding the boundaries of linguistic research. However, traditional analytical methods struggle to cope with the vast and complex multimodal data. This is where novel computational models come into play. Their introduction not only improves analytical efficiency but also uncovers hidden relationships within data through the power of algorithms, broadening our understanding of discourse. This paper, using specific examples, explores how computational models and multimodal discourse analysis complement each other in practice and evolve together. By delving into this process of interactive development, the study aims to provide theoretical support and practical guidance for methodological innovation in multimodal discourse analysis and optimization of computational models, offering effective strategies and technical support for handling increasingly complex multimodal data.

2. The application of computational models in multimodal discourse analysis

2.1. Basic types and functions of computational models

Computational models, as effective tools for processing and analyzing large-scale datasets, play a central role in multimodal discourse analysis. These models primarily include statistical models, machine learning models, and deep learning models, each with distinct functionalities and advantages. Statistical models focus on identifying patterns and structures within data, machine learning models derive decision-making rules through training datasets, and deep learning models use multi-layered network structures to process unstructured big data. In multimodal discourse analysis, these computational models not only efficiently classify, predict, and cluster data but also deeply uncover the intrinsic connections between diverse data modalities such as text, images, and audio [1].

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). https://aei.ewadirect.com

2.2. Automated recognition and analysis of multimodal resources

Multimodal discourse analysis relies on the integrated understanding of various carriers of information, including text, images, and audio. Computational models are indispensable in this process:

2.2.1. Text and image correlation analysis

Using natural language processing (NLP) and computer vision technologies, computational models can identify and parse relevant information in text and images. For example, through image recognition technology, models analyze objects and scenes in images, while NLP tools process descriptive text related to the images. These models establish semantic links between text and images, enabling richer and deeper content understanding [2].

2.2.2. Integrated processing of speech and non-verbal elements

Speech plays a significant role in multimodal analysis. Computational models convert speech content into text via speech recognition technology and use sentiment analysis tools to evaluate speakers' emotions and intentions. Additionally, non-verbal elements such as gestures and facial expressions are incorporated using video analysis technologies, making discourse analysis more comprehensive and precise.

2.2.3. Case study: Applications of computational models in multimodal analysis

Concrete case studies vividly demonstrate the practical utility of computational models in multimodal discourse analysis. For instance, one study analyzed video recordings of political speeches using hybrid models, comprehensively examining the speakers' language, tone, gestures, and facial expressions. This revealed the speakers' persuasive strategies and audience reactions. The case highlights both the ability of computational models to process complex multimodal data and the far-reaching impact of multimodal discourse analysis in practical applications [3].

3. Challenges and impacts of multimodal discourse analysis on computational models

3.1. Challenges posed by the complexity of multimodal data to computational models

The complexity of multimodal data lies primarily in its heterogeneity and unstructured nature, presenting new challenges to computational models. First, different modalities of data (e.g., text, images, and audio) have unique formats and characteristics, requiring distinct processing methods and technologies [4]. For instance, text data needs natural language processing techniques, while image and video data require computer vision support. Moreover, the intrinsic relationships between these data types are often hidden and complex. Accurately identifying and utilizing these connections through computational models is key to enhancing multimodal analysis outcomes.

3.2. Application of multimodal theory in model design

Multimodal discourse analysis theories provide new perspectives and methodologies for the design of computational models. These applications are primarily reflected in:

3.2.1. Adaptive model improvement

Based on multimodal theory, model designers can adapt computational models to better process specific types of multimodal data. For instance, incorporating modal correlation analysis enables models to handle text-image interactions more precisely or synchronize analysis of audio and visual data in videos.

3.2.2. Demand for new algorithm development

As multimodal theory advances, the need for new algorithms grows. These algorithms must be capable of cross-modal data integration and analysis, such as synchronously analyzing text, sound, and image data to extract comprehensive information. Additionally, they must account for cultural and contextual features of multimodal communication, enhancing the models' universality and robustness [5].

4. Interaction mechanisms between computational models and multimodal discourse analysis

4.1. Theoretical framework for mutual adaptation and optimization

The interaction between computational models and multimodal discourse analysis is driven by mutual adaptation and optimization. Computational models continuously learn the characteristics of multimodal data to refine their algorithms, while multimodal discourse analysis theories progressively integrate into the improvement and design of computational models. This bidirectional adaptation enhances computational models' ability to handle complex data and enriches the theoretical foundation and application scope of multimodal discourse analysis.

4.2. Trends in fusion models development

This interaction mechanism has propelled fusion models into the forefront of research and applications. Fusion models are designed to accommodate data inputs from multiple modalities, enabling comprehensive analysis of text, images, and audio. Major development trends include:

4.2.1. Design and implementation of hybrid learning algorithms

Hybrid learning algorithms combine various learning techniques such as supervised, unsupervised, and semi-supervised learning to adapt to the distinct characteristics of different modalities. These algorithms autonomously extract the most effective features from each modality, improving the accuracy and efficiency of models in practical applications [6].

4.2.2. Cross-disciplinary research cases and outcomes

Cross-disciplinary research cases have demonstrated the effectiveness of fusion models in handling complex multimodal scenarios such as social media analysis, educational technology, and health information systems. These cases not only highlight the practical utility of fusion models but also validate their theoretical innovativeness and forward-looking potential.

5. Future research directions and prospects

5.1. Potential impacts of technological advances

With ongoing technological advancements, the development of computational models and multimodal discourse analysis will see expanded prospects. Future breakthroughs, particularly in artificial intelligence and machine learning, are expected to significantly enhance the processing capabilities and analytical precision of computational models. This will enable deeper exploration of the intricate relationships within data, yielding more detailed and comprehensive analytical results. Additionally, with the proliferation of IoT and big data technologies, real-time multimodal data analysis will become feasible, opening new application scenarios for real-time decision support and instant information processing.

5.2. Ethical considerations and practical challenges

As advancements in computational models and multimodal discourse analysis continue, ethical and legal issues must also be addressed. Privacy protection, data security, and ethical use are becoming increasingly critical concerns. Balancing enhanced analytical capabilities with safeguarding personal privacy and data security will be a focal point of future research. Moreover, the diverse and unstructured nature of multimodal data sources presents challenges in data standardization and quality control [7].

5.3. Strategies for integrating computational and multimodal analysis theories

To promote effective integration between computational models and multimodal discourse analysis, future research should focus on developing more general and flexible models capable of meeting specific demands across industries and domains. Furthermore, fostering interdisciplinary collaboration to bridge computational science, linguistics, psychology, and other fields will facilitate comprehensive theoretical and methodological advancements. Establishing robust standards and evaluation systems is also vital for ensuring research quality and driving practical applications [8].

6. Conclusion

This article delves into the interactive development of computational models and multimodal discourse analysis theory, revealing their mutual promotion in practice and theoretical complementarity. Through specific case studies and theoretical analysis, it is evident that computational models significantly enhance the efficiency and depth of multimodal discourse analysis while driving their own innovation and optimization. Faced with increasing data complexity, this integration provides novel perspectives and powerful tools for addressing real-world problems. With continued technological advancements and deeper interdisciplinary collaboration, the combination of computational models and multimodal discourse analysis will further deepen, paving the way for broader applications and in-depth analysis of multimodal data. Addressing ethical and legal concerns alongside technological progress will be crucial for ensuring the sustainable and healthy development of this field.

References

- [1] Kunkun, Z., Emilia, D., & Jane, T. (2022). Evaluating a children's television show as a vehicle for learning about historical artefacts: the value of multimodal discourse analysis. *Discourse: Studies in the Cultural Politics of Education*, 43(6), 866-885.
- [2] Junlong, R. (2021). Study on Automatic Evaluation Method of Spoken English Based on Multimodal Discourse Analysis Theory. *Security* and Communication Networks, 2021.
- [3] Boyue, Z., Guohe, H., Lirong, L., et al. (2021). Development of a multi-factorial enviro-economic analysis model for assessing the interactive effects of combined air pollution control policies. *Resources, Conservation & Recycling*, 175.
- [4] Xiangdan, W. (2020). Construction of Tourism Network Publicity Discourse Based on Multimodal Discourse Analysis Theory. In Liaoning Translation Society, Organizing Committee of the International Academic Forum on Northeast Asian Language, Literature and Translation (Eds.), Proceedings of the 9th International Symposium on Northeast Asian Language, Literature and Translation (pp. 6). Eastern Liaoning University. DOI:10.26914/c.cnkihy.2020.055713
- [5] Triliana, T., & Asih, M. C. E. (2019). The development of the computer-based instructional media with the interactive tutorial model. *Journal of Physics: Conference Series*, 1157(3), 032118-032118.
- [6] Zhang, Y. (2012). The Point-Line-Surface Conceptual Model of the Area Logistics System Based on the Interactive Development of the Manufacturing and Logistics Industry. *Advanced Materials Research*, 1619(452-453), 695-699.
- [7] Kaewkasi, C., & Gurd, R. J. (2007). A distributed dynamic aspect machine for scientific software development. University of Manchester; University of Manchester.
- [8] Jun, S., & Puri, M. V. (2004). DEVELOPMENT OF INTERACTIVE COMPUTATIONAL MODEL OF TEMPERATURE AND MOISTURE DISTRIBUTIONS OF MICROWAVED FOODS. *Applied Engineering in Agriculture*, 20(5), 677.