

# The Important Role of Natural Language Processing in Computational Communication Research

Shuguo Lian

PICC Property and Casualty Co., Ltd. Beijing 100022, China

**Abstract:** *The rapid development of information technology has made the paradigm shift in communication studies increasingly evident, leading to a growing reliance on text data mining techniques in disciplinary research. This satisfies the developmental requirements of computational communication studies. This paper explores natural language processing (NLP) techniques, analyzing their role in integrating these technologies into computational communication research. The following conclusion is drawn: In the context of global information globalization, NLP technology is advantageous in addressing disciplinary issues.*

**Keywords:** Natural Language Processing Technology; Computational Communication; Word Frequency Analysis; Semantic Modeling

## 1. INTRODUCTION

Natural language processing plays a crucial role in computational communication research. It is a core branch of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human natural language [1-3]. In the field of computational communication research Natural language processing promotes the convenience of human-computer interaction [4]. Users can interact with computers using everyday language without the need to learn specific computer languages or commands, greatly reducing the barrier to entry and enabling more people to easily communicate with computers. In computational communication, a large amount of information exists in the form of natural language, such as text, speech, etc [5-6]. Natural language processing technology can automatically analyze this information, extract key content, help users quickly understand the main idea and intention of the information, and make more accurate decisions. In the context of globalization, the demand for cross linguistic communication is increasing day by day [7]. Natural language processing technology can achieve automatic translation and conversion between different languages, break down language barriers, and promote international cooperation and communication. Natural language processing plays an irreplaceable role in computational communication research. It not only enhances the convenience of human-computer interaction, but also improves the efficiency of information extraction and understanding, and promotes the development of cross linguistic communication [8]. With the continuous advancement of technology, natural language processing will play a more important role in the field of computational communication, bringing more convenience and possibilities to human communication [9].

## 2. Overview of Natural Language Processing Technology

The emergence of computer technology has changed the current way of human survival [10]. People can not only enjoy the convenience brought by the internet through computers in their daily lives, but also feel the secure and efficient lifestyle provided by the Internet of Things [11]. At present, in order to meet people's needs for intelligent and modern social development, natural language processing technology can be integrated into current research in computational communication to ensure that machines do not require human management during subsequent operations [12]. They can refer to corresponding control strategies and complete machine function operations on their own, thereby improving the real-time performance of intelligent control work and meeting the requirements of human machine interaction work [13].

### 2.1 Application Overview

Natural language processing technology, as a technical method for achieving language interaction between machines and humans, can complete the conversion of unstructured data such as language and audio, reduce the difficulty of machines understanding human language, and ensure smooth language interaction between machines

and humans [14]. In practical applications, natural language processing technology has two major functions: natural language understanding and natural language generation. Among them, natural language understanding can help machines understand human language, while natural language generation can ensure smooth communication between machines and humans [15]. When standardizing unstructured content, machine learning and deep learning can be used to complete word segmentation, stem extraction, word form restoration, part of speech tagging, named entity recognition, and chunk processing of unstructured content. In the research of computational communication that integrates natural language [16], the processing methods for machine learning natural language mainly include text analysis, text aggregation, correlation analysis, and trend prediction. Among them, text analysis can improve the classification methods of models through supervised learning in the application process [17]. The text clustering method can provide feature vectors and classification books in the application process, and complete the clustering analysis of sample feature vector similarity. The correlation analysis method can identify the relationship between feature vectors and results during the application process. The trend prediction rule can analyze the time series distribution conditions of existing data and complete data prediction work [18].

## 2.2 Fundamentals of Statistical Models

In the current research process of computational propagation, commonly used natural language system models include Bayesian formula conditional probability, N-gram model, HMM model, etc. In specific information analysis work, selecting and using appropriate models based on the type of information and the needs of analysis workload can provide strong support for improving the quality and efficiency of analysis work [19].

### (1) Bayesian formula conditional probability

In the process of applying Bayesian formula conditional probability judgment to determine which set of words  $W$  belongs to  $A_1$  or  $A_2$ , the probabilities  $P(A_1 | W)$  and  $P(A_2 | W)$  can be calculated first, and the values of the two probabilities can be compared. If the probability value of  $A_1$  is greater than that of  $A_2$ , it means that  $W$  belongs to  $A_1$ , otherwise  $W$  belongs to  $A_2$ .

### (2) N-element model

The N-meta model is a commonly used mathematical model in natural language processing. In the process of applying this model, if the natural language meets the Markov property, the probability of the occurrence of a word  $w$  in a sentence  $r$  can be calculated by the formula  $P(w_i | w_1 w_2 \dots w_{i-1}) = P(w_i | w_{i-n+1} \dots w_{i-1})$ , and a formula for calculating the sentence probability can be derived:

$$p(S) = p(w_1 w_2 \dots w_m) = p(w_1) p(w_2 | w_1) \times \dots \times p(w_i | w_{i-n+1} \dots w_{i-1})$$

In the actual analysis process, the larger the value of  $n$  in the formula, the higher the accuracy of the model. However, considering that the parameters used in the model and the required training set will also increase with the increase of  $n$ , and the training set is not large enough, it will lead to a decrease in the number of occurrences of most molecules or conjunctions in the corpus, or even the absence of this situation, which will inevitably lead to data sparsity. Knowledge shortage affects the performance of the model in subsequent processing work, so when applying n-source models for analysis work, data smoothing is required [2].

### (3) HMM model

HMM model is an extremely important mathematical model, which is widely used in natural language processing and speech recognition. In the current research work of computational communication, this model has also been applied to other fields.

## 3. The Application Functions of Natural Language Processing Technology

In the current development process of information technology, the application principle of natural language processing technology is to use preprocessing techniques to standardize unstructured content [20]. The standardized content can be actively and correctly understood and specific functions can be completed. In the current research process of computational communication, natural language processing technology is widely used for functions including word frequency analysis, sentiment analysis, and semantic modeling [21].

### 3.1 Word Frequency Analysis

Word frequency analysis is one of the most commonly used functions in current computational communication science. In the process of ranking analysis of hot words on websites such as Baidu Index and Tiktok Hot List, word frequency analysis technology can help the above websites use word segmentation technology to preprocess the corresponding data information, and then help the distributed big data system stream processing method to complete the quantitative statistics of word information.

### 3.2 Emotion Analysis

Emotion analysis technology is an analysis method that uses algorithms to refer to specific types of emotions to complete text classification processing. Common emotion classifications include positive, negative, happy, sad, and other emotions. In the process of applying sentiment analysis technology, effective analysis of attitudes towards textual information evaluation objects and evaluated subjects, as well as communication topics and emotional tendencies in cyberspace can be achieved through sentiment classification, subjective judgment, opinion summarization, and comment validity analysis. In the current research process of computational communication, sentiment analysis technology is mainly applied in word-of-mouth analysis. In the analysis process, not only can the evaluation of positive or negative information be completed based on the rating level, but also the subjective evaluation sentiment tendency of the evaluators can be analyzed through manual annotation, dictionary matching and other methods, providing reference for the development of product optimization work [3].

### 3.3 Semantic Modeling

Considering the extreme complexity of human language, words have different meanings in different contexts. To improve the accuracy of word analysis, semantic modeling can be introduced into computational communication research, using algorithms combined with context to analyze the hidden meanings behind words. In current computational communication, semantic modeling technology is a relatively mature topic analysis technique. For example, semantic modeling techniques can be applied to extract the semantics of forum network users, compare user knowledge frameworks and explain similarities and differences, and complete research on the impact of response information on audiences.

## 4. The Application Value of Natural Language Processing Technology

### 4.1 Meet the Demands of Computational Communication Research

In the current process of social development, the application of natural language processing technology can effectively meet the demands of computational communication research in communication paradigm adjustment, textual research, and other aspects [22-23].

#### (1) Paradigm adjustment in communication studies

From the perspective of paradigm adjustment in communication studies, the deepening of computational communication research has expanded the study of communication phenomena beyond functional research. Instead, it has begun to describe complex and diverse communication phenomena and group characteristics, and extract new research themes and their underlying meanings during the research process. At the same time, computational communication research can effectively balance the subjective thinking of researchers and the objective requirements of textual research in the application process. With the rapid development of information technology, computational communication research has integrated fragmented, low probability, and large capacity event information into scientific research content, enhancing the depth of communication element research.

#### (2) Text research

At present, there is an urgent need for tools related to textual research in the paradigm of computational communication studies. Specifically, due to the fact that image data is generally a collection of RGB pixel files in matrix format, it is necessary to comprehensively apply knowledge from disciplines such as image processing and computer vision in the process of processing image data. This situation greatly increases the difficulty of research on image data. In the face of the above situation, text has become the main object of current news communication research and the main content of data mining and analysis in computational communication studies due to its

advantages such as low difficulty in obtaining information in cyberspace, less background knowledge required for data processing, and low computational complexity. In the current research process of computational communication, network data acquisition and analysis is one of the important links in the research work. However, in the context of the information age, research in computational communication is facing enormous challenges due to factors such as the lack of regulation over network data sources and exponential growth in data volume. In order to improve the quality and efficiency of research work as much as possible, building an effective set of computational and text semantic recognition methods in the current research process of computational communication, and using them as analysis tools for semantic texts in cyberspace has become an important measure to enhance the data acquisition ability, processing efficiency, and analysis ability of research work. In this context, natural language processing technology has attracted the attention of researchers, and applying this tool to the current research process of computational communication can effectively meet the needs of researchers for personalized and customized data acquisition, organization, and processing [23].

#### **4.2 Create Opportunities for Research in Computational Communication Studies**

The continuous development of natural language processing technology has brought new directions for the research of computational communication, and also provided new ideas for solving problems in traditional computational communication research.

Firstly, the rapid development of natural language reading comprehension technology has propelled the study of computational communication texts towards refinement and customization. Specifically, natural language reading comprehension algorithms can enable machines to complete semantic understanding tasks according to context during application. Summarize, generalize, and refine the key points of language, effectively addressing language diversity, ambiguity, robustness, and knowledge dependence. It should be noted that although natural language reading comprehension technology has significantly improved its technical capabilities after the development of knowledge dependent machine reading comprehension technology, there are still certain difficulties in detecting unanswerable questions and partitioning reasonable answers in the actual text analysis process, which reduces the quality level of research in computational communication.

Secondly, in current research on computational communication, researchers often first propose corresponding models and then use statistical and other disciplinary methods to complete information correlation verification. However, due to the fact that textual data often contains a large number of hidden variables that are difficult to detect and have characteristics such as small sample size, dispersion, and high dimensionality, applying the above methods to carry out text correlation verification will inevitably increase the difficulty of model design verification. To solve this problem, unsupervised learning natural language processing techniques can be applied in the analysis process to complete the analysis of text data feature vectors without providing a specific model in advance, timely identify hidden variables contained in text information, and then carry out model design work based on this, in order to make up for the inherent shortcomings of the assumed model.

Again, in the current research process of computational communication, the application of natural language processing technology can further enhance the objectivity of research work. Specifically, in order to shorten the distance between computational communication and objective science, researchers in computational communication want to find more objective ways to study communication objects as their research focus. In recent years, with the continuous optimization and upgrading of natural language processing technology, its application in computational communication research has become an important measure to improve the accuracy and objectivity of research results. For example, in practical research work, researchers can replace survey questionnaires and interviews with datasets and validation sets under information technology mining. By analyzing massive amounts of data and weakening individual heterogeneity, the reliability and accuracy of data mining work can be improved.

Finally, with the continuous optimization and upgrading of natural language processing technology, research on communication mode analysis and ethical exploration has become increasingly in-depth. Specifically, in practical applications, natural language processing technology not only has language comprehension capabilities but also natural language generation capabilities, which can be applied to the generation of data news. In recent years, with the continuous development of natural language processing technology, this technology not only effectively meets the current demand for actively generating text in the cyberspace, but also provides new topics for the research of computational communication research objects, models, and ethics.

## 5. Conclusion

In summary, computational sociology can effectively analyze complex social systems by collecting and analyzing data related to human behavior in cyberspace, eliminating heterogeneity and noise factors, and enriching people's social cognition while identifying social phenomena in the system. At present, conducting research on the application value of natural language processing in computational communication can provide assistance for the development of computational sociology.

## References

- [1] Collobert, R. , Weston, J. , Bottou, Léon, Karlen, M. , Kavukcuoglu, K. , & Kuksa, P. . (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(1), 2493-2537.
- [2] Zadeh, L. A. . (2002). Precisiated natural language - toward a radical enlargement of the role of natural languages in information processing, decision and control. *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on. IEEE*.
- [3] Joshi, A. K. . (1998). Role of constrained computational systems in natural language processing. *Artificial Intelligence*, 103(1-2), 117-132.
- [4] Shams, R. . (2014). Identification of informativeness in text using natural language stylometry.
- [5] Juliet, Beni, Edgcomb, Bonnie, & Zima. (2019). Machine learning, natural language processing, and the electronic health record: innovations in mental health services research. *Psychiatric services (Washington, D.C.)*.
- [6] Wang, Z., Yan, H., Wang, Y., Xu, Z., Wang, Z., & Wu, Z. (2024). Research on autonomous robots navigation based on reinforcement learning. *arXiv preprint arXiv:2407.02539*.
- [7] Zhu, Z., Wang, Z., Wu, Z., Zhang, Y., & Bo, S. (2024). Adversarial for Sequential Recommendation Walking in the Multi-Latent Space. *Applied Science and Biotechnology Journal for Advanced Research*, 3(4), 1-9.
- [8] Shen, Z. (2023). Algorithm Optimization and Performance Improvement of Data Visualization Analysis Platform based on Artificial Intelligence. *Frontiers in Computing and Intelligent Systems*, 5(3), 14-17.
- [9] Yao, J. (2024). The Impact of Large Interest Rate Differentials between China and the US on the Role of Chinese Monetary Policy -- Based on Data Model Analysis. *Frontiers in Economics and Management*, 5(8), 243-251.
- [10] He, C., Yu, B., Liu, M., Guo, L., Tian, L., & Huang, J. (2024). Utilizing Large Language Models to Illustrate Constraints for Construction Planning. *Buildings*, 14(8), 2511. <https://doi.org/https://doi.org/10.3390/buildings14082511>
- [11] Yang, H., Zi, Y., Qin, H., Zheng, H., & Hu, Y. (2024). Advancing Emotional Analysis with Large Language Models. *Journal of Computer Science and Software Applications*, 4(3), 8-15.
- [12] Zheng Ren, "Balancing role contributions: a novel approach for role-oriented dialogue summarization," *Proc. SPIE 13259, International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2024)*, 1325920 (4 September 2024); <https://doi.org/10.1117/12.3039616>
- [13] Liu, S., Li, X., & He, C. (2021). Study on dynamic influence of passenger flow on intelligent bus travel service model. *Transport*, 36(1), 25 - 37. <https://doi.org/10.3846/transport.2021.14343>
- [14] Zhu, Z., Wang, Z., Wu, Z., Zhang, Y., & Bo, S. (2024). Adversarial for Sequential Recommendation Walking in the Multi-Latent Space. *Applied Science and Biotechnology Journal for Advanced Research*, 3(4), 1-9.
- [15] He, C., Liu, M., Zhang, Y., Wang, Z., Hsiang, S. M., Chen, G., Li, W., & Dai, G. (2023). Space - Time - Workforce Visualization and Conditional Capacity Synthesis in Uncertainty. *Journal of Management in Engineering*, 39(2), 04022071. <https://doi.org/10.1061/JMENEA.MEENG-4991>
- [16] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. *Journal of Theory and Practice of Engineering Science*, 4(02), 15 - 22. [https://doi.org/10.53469/jtpes.2024.04\(02\).03](https://doi.org/10.53469/jtpes.2024.04(02).03)
- [17] Bonilla, M., Rasdorf, W., Liu, M., Al-Ghandour, M., & He, C. (2023). Inequity Reduction in Road Maintenance Funding for Municipalities. *Public Works Management & Policy*, 28(3), 339 - 362. <https://doi.org/10.1177/1087724X221130047>
- [18] Zheng, H., Wang, B., Xiao, M., Qin, H., Wu, Z., & Tan, L. (2024). Adaptive Friction in Deep Learning: Enhancing Optimizers with Sigmoid and Tanh Function. *arXiv preprint arXiv:2408.11839*.
- [19] Ren, Z. (2024). A Novel Topic Segmentation Approach for Enhanced Dialogue Summarization. *World Journal of Innovation and Modern Technology*, 7(4), 42-49.

- 
- [20] He, C., Liu, M., Wang, Z., Chen, G., Zhang, Y., & Hsiang, S. M. (2022). Facilitating Smart Contract in Project Scheduling under Uncertainty — A Choquet Integral Approach. *Construction Research Congress 2022*, 930 - 939. <https://doi.org/10.1061/9780784483961.097>
- [21] Wang, Z., Zhu, Y., He, S., Yan, H., & Zhu, Z. (2024). LLM for Sentiment Analysis in E-commerce: A Deep Dive into Customer Feedback. *Applied Science and Engineering Journal for Advanced Research*, 3(4), 8-13.
- [22] Ji, H., Xu, X., Su, G., Wang, J., & Wang, Y. (2024). Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising. *Academic Journal of Science and Technology*, 9(2), 215-220.
- [23] Tian, Q., Wang, Z., Cui, X. Improved Unet brain tumor image segmentation based on GSConv module and ECA attention mechanism. *arXiv preprint arXiv:2409.13626*.