Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.2026020814

Analysis of Semantic Disambiguation Techniques in Dialogue Systems

Xinyue Liu*

Department of Intelligent Science and Technology, Beijing University of Posts and Telecommunications, Beijing, China

**Corresponding author: Xinyue Liu*

Abstract

With the rapid advancement of Artificial Intelligence (AI), human-computer dialogue systems have emerged as a prominent application in the AI field, where semantic disambiguation plays a critical role. This paper provides a comprehensive survey and analysis of semantic disambiguation techniques in Natural Language Processing. It not only examines the strengths and limitations of traditional approaches but also investigates state-of-the-art methods that are shaping the field. Through a systematic review, this study outlines the overall development trajectory of disambiguation technologies and identifies key challenges that remain unresolved. While significant breakthroughs have been achieved in semantic disambiguation, this review reveals that substantial challenges persist in areas such as contextual understanding, cross-domain adaptation, and real-time processing. Overall, this paper emphasizes the growing importance of integrating knowledge-driven and data-driven approaches, highlighting the potential of hybrid models that combine linguistic rules, contextual embeddings, and large language models. The findings aim to offer valuable insights for future research directions and practical implementations in evolving dialogue systems.

Keywords

human-computer dialogue systems, artificial intelligence, word sense disambiguation, natural language processing

1. Introduction

Dialogue System refers to a computer system that realizes human-computer interaction through natural language, and its core goal is to simulate human's ability to have a conversation. Ambiguity is a pervasive phenomenon in language which occurs at all levels of linguistic analysis [1]. However, based on the in-depth analysis of the massive amount of actual communication data, it can be found that there are still a large number of urgent language comprehension problems in the communication between humans and AI. These problems not only directly hinder the smoothness of human-machine collaboration, but also become the key bottleneck that restricts the in-depth penetration of AI technology into the field of production -- when there is a bias in the semantic transmission, it is difficult for AI to play a precise auxiliary role in the production process, and the release of its technological value is also limited.

Traditional techniques like Naive Bayes and Hidden Markov Models are applied to many NLP) disambiguation tasks such as Part-of-Speech tagging, Word Sense disambiguation and Text Categorization etc [2]. The introduction of Transformer-based models significantly improved ambiguity resolution. Recent studies have integrated multiple approaches to enhance Word Sense Disambiguation (WSD). For instance, Gangadharan et al. identified fastText as the most effective word embedding technique for sense representation [3]. Dhanashree et al. proposed a graph-based Lesk approach for Hindi WSD [4]. Kavitha et al. further combined mutual information with neural classifiers by using supervised learning to improve English WSD accuracy [5]. Despite progress, significant challenges remain in the systematic understanding and resolution of semantic ambiguity in AI-driven dialogue systems. IEEE Std 3128-2025 reveals that persistent deficiencies in WSD, such as contextual misunderstanding and poor generalization, remain a major obstacle to reliable AI dialogue systems [6]. Based on the systematic evaluation by Yang et al., large multimodal models demonstrate limited ability in scrutinizing and disambiguating faulty or ambiguous textual inputs, particularly when semantic conflicts arise between different modalities [7]. Addressing this gap is critical for improving interactive experience, guiding model optimization, and enabling reliable applications in specialized domains such as healthcare and finance.

This study aims to provide a summary by systematically evaluating the disambiguation capabilities of cutting-edge technologies in 2025, benchmarking them against traditional methods to quantify their advancements and limitations. The primary objectives are: 1) to assess the technical value and efficacy of state-of-the-art WSD models; 2) to identify and analyze the persistent challenges and failure modes in real-world scenarios; and 3) to explore viable pathways and directions for future algorithm optimization, thereby facilitating the robust application of AI dialogue systems in critical domains such as medical and financial services.

2. Methods

2.1 Traditional Techniques

2.1.1 Naïve Bayes Classifier

The Naïve Bayes classifier is a simple probabilistic classification algorithm based on Bayes' theorem, widely applied in various NLP disambiguation tasks such as part-of-speech tagging, word sense disambiguation, and text classification [2]. The disambiguation process generally consists of four steps: First, identifying the ambiguous word within a sentence and determining its possible sense categories, while treating surrounding words and their part-of-speech tags as contextual features. Second, applying Bayes' theorem to compute the probability of each sense category given the observed contextual features.

The formula is as follows:

$$P(A|B) \quad = \frac{P(B|A)}{P(B)} \cdot P(A) \tag{1}$$

Where A represents a sense category and B denotes the contextual feature set. The sense category with the highest probability is selected as the final disambiguation result. As a foundational technique, the Naïve Bayes classifier has been extended and refined by numerous researchers through integration with other methods. For instance, He et al. developed a feature auto-selection based Naïve Bayes model for Chinese word sense disambiguation [8]. Similarly, Zhang et al. incorporated syntactic and part-of-speech features extracted from parse trees, using the Naïve Bayes classifier trained on sense-annotated corpora to disambiguate words in test data [9]. Owing to its dependence on simple probabilistic multiplications, the method exhibits high computational efficiency, making it suitable for real-time interactive applications and scenarios with limited training data. However, it also suffers from limitations such as the strong feature independence assumption and inaccurate sense prediction for rare words.

2.1.2 Hidden Markov Model (HMM)

HMM is a statistical model used to describe a Markov process with unobserved hidden states. By modeling the probabilistic relationship between hidden states and observed sequences, HMM is well-suited for

processing temporal and sequential data. The disambiguation process based on HMM can be divided into four steps: First, defining the hidden states (e.g., word senses) and the observation sequence (e.g., words or contextual features). Second, estimating the three core parameters—initial probability, transition probability, and emission probability—from annotated corpora. Third, decoding the observation sequence using the Viterbi algorithm. Finally, the optimal path computed by the Viterbi algorithm is output as the disambiguation result, assigning the most probable sense to each word. Building upon this foundational technique, researchers have continued to develop more powerful disambiguation models. For example, Li et al.integrated HMM, Maximum Entropy (MaxEnt), and Conditional Random Fields (CRF) with rules derived from improved mutual information, increasing the average part-of-speech tagging accuracy by 5% [10]. More recently, Weiming Shao et al.proposed a Semi-Supervised Robust Hidden Markov Regression (SsRHMR) model with a distributional learning algorithm to enhance HMM performance in scenarios with scarce labeled samples and data anomalies [11]. Compared to the Naïve Bayes classifier, HMM exhibits a stronger capability in handling long-distance contextual dependencies, making it more suitable for human–machine dialogue systems. Nevertheless, it still faces challenges in processing very long texts and accurately interpreting low-frequency word senses.

2.1.3 Transformer

Transformer architecture is a neural network architecture based on the self-attention mechanism, first introduced by the Google team in 2017. It has fundamentally transformed the field of NLP and serves as the core foundation of modern large language models such as BERT and GPT. Transformer achieves dynamic semantic disambiguation through its self-attention mechanism. First, the model encodes the entire sentence into a sequence of context-aware vector representations, where the embedding of each word incorporates global contextual information. Then, for the output vector corresponding to the ambiguous word in the final layer, sense determination is performed via one of the following two approaches: (1) employing a classification layer to directly predict a predefined sense label, or (2) calculating its similarity with sense definition vectors from a lexical database and selecting the most matching sense.Building upon this architecture, researchers have explored various methods to improve disambiguation accuracy. For instance, Gregor et al. investigated the use of contextualized embeddings from BERT for interpretable word sense disambiguation [12]. Edoardo et al. reframed WSD as a span extraction task leveraging Transformer's powerful contextual understanding, proposing the ESC model and introducing gloss noise to mitigate data bias [13]. Compared to traditional methods, Transformer-based models demonstrate a stronger capability to capture nuanced contextual information and exhibit superior generalization abilities, largely due to large-scale pre-training. However, they also come with drawbacks such as high computational costs, lower inference efficiency, and significant model complexity.

2.2 Emerging Techniques

2.2.1 Adaptive Semantics Learning Strategy with Reward–Penalty Mechanism

Li et al.[14]Li et al.[14]Li et al. (2025) proposed an adaptive learning method incorporating a reward–penalty mechanism to balance the semantic distribution in polysemous triggers and arguments. This framework consists of four main modules: the reward–penalty mechanism, the SESA mechanism, a semantics-enhanced encoder, and a task decoder. The reward–penalty mechanism dynamically adjusts the learning process for polysemous words by reinforcing correctly learned semantics while penalizing incorrect ones based on semantic probability distributions and model classification results. The SESA mechanism generates accurate and comprehensive representations for all event mentions within a sentence. The semantics-enhanced encoder encodes tokens into vector representations and enriches their semantics with information from all event mentions. The task decoder identifies all potential triggers and argument candidates in the sentence and classifies their types. The main operational procedure follows a sequential pipeline: it begins with trigger encoding, followed by trigger decoding, then proceeds to argument encoding, and finally concludes with argument decoding. Experimental results demonstrate that this approach exhibits strong scalability and generalization capability. Compared to traditional methods, it dynamically adapts to the context-dependent meanings of polysemous words, prevents the model from relying on superficial or local statistical features in the data, and achieves superior performance.

2.2.2 Fast2Vec

In human-computer dialogue, while most users engage only at the level of casual conversation, the application of such dialogue systems in specialized domains requires AI to possess the ability to analyze and comprehend complex professional texts. To address this, Ayu Pertiwi et al. integrated the semantic advantages of Word2Vec with the subword modeling capability of FastText to design Fast2Vec [15]. Its core innovations lie in subword embedding, Dynamic Topic Modeling (DTM) integration, and context-sensitive semantic similarity computation.Fast2Vec adopts the character-level n-gram representation from FastText, which offers distinct advantages in semantic disambiguation, including handling Out-of-Vocabulary (OOV) words and capturing morphological information. In terms of DTM integration, Fast2Vec incorporates UMAP for dimensionality reduction, Affinity Propagation (AP) for clustering, and enhanced lexical representations to achieve more coherent dynamic topic analysis. For context-aware semantic similarity measurement, unlike traditional FastText which averages subword vectors directly, Fast2Vec introduces a weighted subword contribution mechanism.Comparative experiments involving Word2Vec, FastText, and Fast2Vec demonstrate that Fast2Vec achieves higher semantic similarity scores within relevant topics. Higher scores indicate more stable semantic clustering, which facilitates accurate semantic understanding and reduces ambiguity. Compared to conventional techniques, Fast2Vec effectively handles OOV words, captures evolving semantic shifts, and sensitively detects topic activity across various domains over time, thereby supporting the analysis of current societal development trends.

2.2.3 CLIP-Driven Transformer

In human-AI dialogue, the use of multimodal information carriers such as images can enhance communication and even help resolve semantic ambiguity through the integration of visual and textual cues. Chen et al. proposed a novel CLIP-Driven Transformer architecture capable of learning category-aware representations for precise object localization [16]. The model operates through the following mechanism: an input image is first processed by a Transformer encoder utilizing self-attention to generate feature maps. Simultaneously, a category-aware self-attention mechanism produces attention maps and predicts class labels. The category-aware information is then integrated with the feature maps to generate a localization map, which identifies the positions of specific categories within the image. The overall architecture comprises separate branches for image and text processing, enabling multimodal fusion for semantic understanding and object localization. Compared to conventional methods, this approach demonstrates strengths in multimodal integration, exhibits strong generalization under weak supervision, and shows adaptability to multi-turn dialogue contexts. However, studies indicate that the model underperforms in complex environments where objects are easily camouflaged or visually ambiguous.

3. Discussion

3.1 Challenges

3.1.1 Knowledge Scarcity and Imitation Limitations in Professional Domain Dialogue

Despite the superiority of Transformer and HMM over earlier models in capturing longer contexts, multi-turn dialogues still suffer from historical information decay and long-range dependency issues. Neural dialogue systems rely heavily on large annotated datasets and structured knowledge, learning primarily through imitation of human responses, which often lack diversity—especially in specialized domains. This restricts their ability to conduct deep, professional dialogues [17]. The MMLU team introduced the “Human Last Exam” (HLE) benchmark, comprising 2,700 expert-curated, multimodal questions that resist internet retrieval and are validated against state-of-the-art LLMs [18]. Preliminary results show current models perform poorly on HLE, highlighting that imitation learning alone is insufficient for knowledge-intensive and expert-level conversations.

3.1.2 Noise Interference and Shallow Interaction in Multimodal Fusion

Although multimodal models such as CLIP-Driven Transformer [16] offers new pathways for WSD, several critical challenges remain. Firstly, the visual grounding capability of these models remains unstable in complex real-world environments, where irrelevant background information often interferes with model

reasoning, leading to degraded performance in vision-assisted WSD. Secondly, current multimodal integration methods mostly rely on simple feature concatenation or alignment, rather than achieving deep semantic collaboration between visual and textual information. The lack of organic interaction and complementarity between image content and linguistic context limits the full potential of multimodal approaches in WSD. Future work must focus on developing more discriminative fusion mechanisms that allow visual signals to participate in language understanding in a more precise and structured manner.

3.1.3 Bottlenecks in Cross-Context and User-Aware Personalized Disambiguation

Current dialogue systems face significant challenges in personalized semantic disambiguation across temporal, user-specific, and linguistic dimensions. Although adaptive mechanisms such as Adaptive Semantics Learning Strategy with Reward–penalty Mechanism [14] have been proposed to handle polysemous words, these approaches remain constrained by static training data and fixed architectures, struggling to adapt to dynamically evolving semantics across time and user groups. Cross-lingual deficiencies are particularly prominent—a joint study by Harvard, MIT, and Microsoft Research evaluated 14 mainstream models (including Llama and Gemini) and revealed widespread weaknesses in cross-language understanding. These limitations not only hinder natural human-AI interaction but also risk widening information disparities among linguistic groups. Achieving truly inclusive conversational AI requires significant improvements in systems' adaptability to multi-dimensional contextual variations.

3.2 Future Prospects

3.2.1 Developing Efficient Long-Range Context Modeling Mechanisms

To address the challenges of information decay and long-range dependencies in multi-turn dialogues, it is essential to explore more efficient architectures beyond standard self-attention, such as linear-complexity attention variants, state space models, or explicit memory augmentation modules. Furthermore, models should advance beyond mere data imitation toward deep semantic reasoning and situational understanding, enabling them to capture complex logical relationships and long-term dependencies in professional texts. This will significantly enhance comprehension and generation quality in specialized domains.

3.2.2 Advancing Deep Multimodal Fusion Learning

There is a need to develop more robust and interpretable multimodal fusion methods that integrate visual, auditory, and other modal inputs to achieve cross-modal semantic alignment and joint reasoning. Such technologies hold promise for breakthroughs in fields like medical diagnosis and intelligent education—for instance, through vision-language modeling for symptom analysis and telemedicine, or via multimodal context understanding to enhance the practicality and inclusivity of interactive devices.

3.2.3 Building Dynamic and Adaptive Personalized Learning Frameworks

It is crucial to advance online learning, continual learning, and reinforcement learning mechanisms based on user feedback, allowing models to continuously adapt to semantic evolution and user preferences during operation and support personalized interactions across languages and cultures. This will enable lifelong learning in dialogue systems, markedly improving their utility in low-resource languages and specialized domains, while reducing dependence on large-scale annotated data.

4. Conclusion

This paper has provided a concise survey and analysis of key techniques in WSD, ranging from traditional classifiers such as Naïve Bayes to cutting-edge approaches like the Adaptive Semantics Learning Strategy. The analysis reveals that persistent challenges remain in areas including contextual understanding, multimodal fusion, and cross-domain adaptability and personalization. It is important to note that the literature reviewed in this study does not encompass the entire field. In future work, the further study will expand the scope of this research by incorporating a wider range of publications and continuously integrating the latest findings to keep the study current and comprehensive.

References

[1] Piantadosi, S. T., Tily, H. and Gibson, E. The communicative function of ambiguity in language. Cognition. 2012, 122(3), pp. 280-291. https://doi.org/10.1016/j.cognition.2011.10.004.

[2] Sharma, H. Improving natural language processing tasks by using machine learning techniques. In 2021 5th international conference on information systems and computer networks (ISCON), Mathura, India, 2021; pp. 1-5. https://doi.org/10.1109/ISCON52037.2021.9702447.

[3] Gangadharan, V. and Gupta, D. Paraphrase detection using deep neural network based word embedding techniques. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020; pp. 517-521. https://doi.org/10.1109/ICOEI48184.2020.9142877.

[4] Kulkarni, D. S. and Rodd, S. F. Word Sense Disambiguation for Lexicon-based Sentiment Analysis in Hindi. Webology. 2022, 19(1), pp. 592-600. https://doi.org/10.14704/WEB/V19I1/WEB19042.

[5] Kavitha, K., Pranav, S. and Anil, A. Word Sense Disambiguation Using Supervised Learning. In 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023; pp. 1-6. https://doi.org/10.1109/GCAT59970.2023.10353351.

[6] IEEE. IEEE Recommended Practice for the Evaluation of Artificial Intelligence (AI) Dialogue System Capabilities: IEEE Std 3128-2025. Piscataway, NJ: IEEE, 2025.

[7] Yang, H., Li, J., Li, G., Chang, Y. and Wu, Y. Can large multimodal models actively recognize faulty inputs? a systematic evaluation framework of their input scrutiny ability. arXiv preprint arXiv:2508.04017. 2025. https://doi.org/10.48550/arXiv.2508.04017.

[8] He, J. Z. and Wang, H. F. Chinese word sense disambiguation based on maximum entropy model with feature selection. Journal of Software. 2010, 21(6), pp. 1287-1295. https://doi.org/10.3724/SP.J.1001.2010.03591.

[9] Zhang, C. X., Luan, B., Gao, X. Y. and Lu, Z. M. Chinese word sense disambiguation based on parsing analysis. Application Research of Computers/Jisuanji Yingyong Yanjiu. 2014, 31(1), pp. 40-47. https://doi.org/10.3969/j.issn.1001-3695.2014.01.008.

[10] Li, H. D., Jia, Z., Yin, H. F. and Yang, Y. Rule-based tagging method of Chinese ambiguity words. Journal of Computer Applications. 2014, 34(8), pp. 2197-2201. https://doi.org/10.11772/j.issn.1001-9081.2014.08.2197.

[11] Shao, W., Han, W., Xiao, C., Chen, L., Yu, M.-Q. and Chen, J. Semi-supervised robust hidden Markov regression for large-scale time-series industrial data analytics and its applications to soft sensing. IEEE Transactions on Automation Science and Engineering. 2024, 22, pp. 5143-5157. https://doi.org/10.1109/TASE.2024.3417019.

[12] Wiedemann, G., Remus, S., Chawla, A. and Biemann, C. Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings. arXiv preprint arXiv:1909.10430. 2019. https://doi.org/10.48550/arXiv.1909.10430.

[13] Barba, E., Pasini, T. and Navigli, R. ESC: Redesigning WSD with extractive sense comprehension. In Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies, online, 2021; pp. 4661-4672. https://doi.org/10.18653/v1/2021.naacl-main.371.

[14] Li, H., Tian, Z., Wang, X., Zhou, Y., Pan, S., Zhou, J., Xu, Q. and Li, D. Handling polysemous triggers and arguments in event extraction: an adaptive semantics learning strategy with reward–penalty

mechanism. Frontiers of Information Technology & Electronic Engineering. 2025, 26(4), pp. 534-555. https://doi.org/10.1631/FITEE.2400220.

[15] Pertiwi, A., Azhari, A. and Mulyana, S. Fast2Vec, a modified model of FastText that enhances semantic analysis in topic evolution. PeerJ Computer Science. 2025, 11, p. e2862. https://doi.org/10.7717/peerj-cs.2862.

[16] Chen, Z., Shen, Y., Cao, L., Zhang, S. and Ji, R. CLIP-driven transformer for weakly supervised object localization. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025, 47(6), pp. 4878-4896. https://doi.org/10.1109/TPAMI.2025.3548704.

[17] Chen, H., Liu, X., Yin, D. and Tang, J. A survey on dialogue systems: Recent advances and new frontiers. Acm Sigkdd Explorations Newsletter. 2017, 19(2), pp. 25-35. https://doi.org/10.1145/3166054.3166058.

[18] Stanford HAI. Artificial Intelligence Index Report 2025. Stanford, CA: Stanford University, 2025.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access