Literature Survey on "N-Gram Language Models in Modern NLP"

Course: Machine Learning and Applications

Group 21

Group Members: Shaojie Zhang

Abstract

This literature survey explores the role of N-gram language models in modern natural language processing (NLP). Despite the dominance of deep learning methods, N-gram models remain relevant due to their simplicity, efficiency, and interpretability. The report examines their evolution, integration with neural architectures, and practical use in resource-constrained environments. It analyzes five recent papers to compare methodologies, performance, and hybrid model designs. Critical insights and future research directions are discussed.

1. Introduction

N-gram language models have historically played a foundational role in NLP by statistically modeling sequences of words. They have been widely used in applications such as machine translation, speech recognition, and spelling correction. While deep learning has transformed NLP, N-gram models continue to be useful, especially in low-resource scenarios. This study aims to assess the continued relevance of N-gram models by reviewing recent research, identifying their strengths and limitations, and evaluating hybrid models that combine N-gram features with neural architectures.

2. Methodology

The literature was sourced from databases including IEEE Xplore, arXiv, and Google Scholar. Keywords such as "n-gram language model", "hybrid NLP models", and "transformers with n-gram" were used. Papers published within the last five years and with significant citations or novel methodologies were prioritized. Five key papers were selected based on their relevance, innovation, and impact on the field.

3. Literature Review

3.1 Thematic Analysis

Recent papers have explored integrating N-gram models into transformer architectures (e.g., N-Grammer), scaling N-gram models (e.g., Infini-Gram), and combining residual learning with N-gram approaches (e.g., N-Gram Is Back). Another line of work includes explicitly modeling phrase-level semantics using N-gram masking (e.g., ERNIE-Gram) and approximating large pre-trained models via N-gram estimation.

3.2 Comparative Analysis

Compared to deep learning models, N-gram approaches are interpretable and computationally efficient. However, they lack the ability to model long-range dependencies. Hybrid approaches leverage the strengths of both paradigms, showing notable improvements in text classification, summarization, and speech recognition tasks.

4. Critical Analysis

The selected studies are well-conducted, with rigorous experiments and clear conclusions. However, many papers focus on benchmark datasets and overlook real-world deployment constraints. While hybrid models show promise, their added complexity and training costs are non-trivial. Research gaps include the application of N-gram models to newer domains such as conversational agents or low-resource languages.

5. Conclusion

This survey reaffirms the value of N-gram models in contemporary NLP. While deep neural networks dominate the landscape, N-gram models remain relevant, especially in scenarios demanding interpretability and efficiency. Future research should focus on optimizing hybrid models, exploring domain-specific applications, and enhancing scalability.

References

[1] Roy et al., 'N-Grammer: Augmenting Transformers with Latent N-Grams', 2022.

[2] Liu et al., 'Infini-Gram: Scaling Unbounded N-Gram Language Models to a Trillion Tokens', 2024.

[3] Li et al., 'N-Gram Is Back: Residual Learning of Neural Text Generation with N-Gram Language Models', 2022.

[4] Xiao et al., 'ERNIE-Gram: Pre-Training with Explicitly N-Gram Masked Language Modeling', 2020.

[5] Anonymous, 'On the N-Gram Approximation of Pre-Trained Language Models', 2023.

Appendix

GitHub Repository Link

The GitHub repository for this project is available at:
https://github.com/ZhangShaoj-NEU/NGram-NLP.git

Data Used

This project is a literature survey and does not use any dataset.
All the references and resources used in this project are listed in the References section.