ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

RICH SHORT TEXT CONVERSION USING SEMANTIC KEY CONTROLLED SEQUENCE GENERATION

[1]SADANALA MANJUSHA, [2]P.BOBBY SOWJANYA

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

Short text data such as search queries, social media posts, and user inputs often lack sufficient context and semantic richness, making it difficult for traditional natural language processing (NLP) systems to generate meaningful outputs. These limitations affect applications such as chatbots, recommendation systems, and content generation. This project proposes a novel approach for transforming short text into rich, context-aware content using semantic key-controlled sequence generation, leveraging advanced deep learning techniques. The proposed system extracts key semantic features from short input text using embedding techniques such as Word2Vec, GloVe, or transformer-based models. These semantic keys act as control signals that guide the sequence generation process. A deep learning architecture, such as a Sequence-to-Sequence (Seq2Seq) model with attention mechanism or transformer-based models like GPT, is employed to generate expanded and meaningful text. The model learns contextual relationships and generates coherent sentences by conditioning on the extracted semantic keys, ensuring that the output remains relevant to the input. The methodology includes preprocessing steps such as tokenization, stop-word removal, and embedding generation. The model is trained on large text corpora to learn language patterns and semantic relationships. During inference, the system takes short input text, extracts semantic keys, and generates enriched output sequences. Experimental results demonstrate that the proposed approach significantly improves text quality, coherence, and contextual relevance compared to traditional text generation methods. However, challenges such as controlling output diversity and computational complexity remain.

Keywords: Short Text Processing, Semantic Key Extraction, Sequence Generation, Natural Language Processing (NLP), Deep Learning, Seq2Seq Model, Transformer, Text Generation, Language Models, Context Enrichment

I.INTRODUCTION

In recent years, the rapid growth of digital communication platforms has led to an increase in the use of short text formats such as search queries, tweets, chat messages, and user inputs. These short texts are often limited in length and lack sufficient contextual information, making it difficult for natural language processing (NLP) systems to fully understand their meaning. As a result, tasks such as text generation, summarization, and recommendation become challenging due to ambiguity and insufficient semantic content. Traditional text processing techniques rely on keyword matching or basic statistical models, which fail to capture deeper contextual relationships within short text.

With the advancement of deep learning and NLP, more sophisticated approaches have been developed to address these challenges. Techniques such as word embeddings (Word2Vec, GloVe) and transformer-based models have enabled machines to understand semantic relationships between words. Sequence generation models, particularly Sequence-to-Sequence (Seq2Seq) architectures and transformer models like GPT, have shown remarkable performance in generating coherent and context-aware text. However, these models often struggle when the input text is too short or lacks sufficient semantic cues, leading to generic or irrelevant outputs.

The proposed system introduces a semantic key-controlled sequence generation approach to enhance short text into rich and meaningful content. The system first extracts important semantic features (keys) from the input text, which serve as guiding signals for the generation process. These semantic keys are then used to control the output of a deep learning model, ensuring that the generated text remains relevant and contextually rich. This approach improves the quality, coherence, and informativeness of generated text. The system can be applied in various domains such as chatbots, content generation, search engines, and intelligent assistants, providing a powerful solution for short text understanding and expansion.

II SURVEY OF RESEARCH

The study by T. Mikolov et al. (2013) [1] introduced Word2Vec, a neural network-based model for learning word embeddings. The methodology focuses on representing words in a continuous vector space where semantic relationships are preserved. Results show that Word2Vec effectively captures word similarities and analogies. However, it lacks contextual understanding of words in different sentences. This research is relevant as it provides a foundation for semantic key extraction in short text processing.

The work by J. Pennington et al. (2014) [2] proposed GloVe (Global Vectors for Word

Representation), which combines global statistical information with local context. The methodology uses word co-occurrence matrices to generate embeddings. Results demonstrate improved semantic representation compared to traditional models. However, it does not handle dynamic context. This study supports the use of embedding techniques for semantic understanding.

The research by I. Sutskever et al. (2014) [3] introduced Sequence-to-Sequence (Seq2Seq) models for text generation. The methodology uses encoder-decoder architectures to map input sequences to output sequences. Results show that Seq2Seq models are effective for translation and text generation tasks. However, they struggle with long sequences and context retention. This research forms the basis for sequence generation in the proposed system.

The study by D. Bahdanau et al. (2015) [4] introduced the attention mechanism to improve Seq2Seq models. The methodology allows the model to focus on relevant parts of the input sequence during generation. Results demonstrate significant improvement in translation and text generation tasks. However, computational complexity increases. This research is important for enhancing context awareness in sequence generation.

The work by A. Vaswani et al. (2017) introduced the Transformer architecture [5], which eliminates recurrence and uses self-attention mechanisms. The methodology enables parallel processing and captures long-range dependencies in text. Results show state-of-the-art performance in various NLP tasks. However, it requires large datasets and computational resources. This research supports the use of transformer-based models in the proposed system.

The research by A. Radford et al. (2019) introduced Generative Pre-trained Transformer (GPT) models [6], which are capable of generating high-quality and context-aware text. The methodology involves pre-training on large corpora and fine-tuning for specific tasks. Results demonstrate superior performance in text generation tasks. However, controlling output relevance remains a challenge. This study is directly related to semantic key-controlled text generation.

Additionally, recent advancements in controllable text generation have explored techniques such as conditional text generation and prompt-based learning. These approaches allow models to generate text based on specific conditions or inputs, improving relevance and coherence. Integrating semantic key control with these techniques further enhances the ability to generate meaningful and context-rich text from short inputs, making the proposed system more effective and adaptable.

III. WORKING METHODOLOGY

The proposed system for Rich Short Text Conversion using Semantic Key-Controlled Sequence Generation begins with input text acquisition and preprocessing. Short text inputs such as queries, chat messages, or short descriptions are collected and processed through techniques such as tokenization, stop-word removal, and normalization. Since short texts often lack sufficient context, the system applies semantic enrichment techniques to extract meaningful features. Word embeddings such as Word2Vec, GloVe, or contextual embeddings from transformer models are used to convert text into vector representations. From these representations, important semantic keys (keywords or concepts) are identified using techniques such as attention mechanisms or keyword extraction algorithms.

In the next stage, the extracted semantic keys are used to control the sequence generation process. A deep learning model, such as a Sequence-to-Sequence (Seq2Seq) model with attention or a transformer-based model like GPT, is trained on large text datasets. The semantic keys act as guiding signals that influence the generation of output sequences. The model learns to generate context-rich and meaningful text by conditioning on both the input text and the extracted semantic features. This approach ensures that the generated content is not only coherent but also relevant to the original short text. Feature fusion techniques may also be applied to combine semantic keys with contextual embeddings for improved performance.

In the final stage, the system generates enriched text output and evaluates its quality. When a new short text input is provided, the system extracts semantic keys and feeds them into the trained model to generate expanded content. The output is evaluated using metrics such as BLEU score, ROUGE score, and semantic similarity measures. The system may also include post-processing steps to improve readability and remove redundant content. This methodology ensures accurate, context-aware, and meaningful text generation, making the system suitable for applications such as chatbots, content generation, and intelligent text assistance systems.

IV RESULTS EXPLANATIONS

The performance of the proposed Rich Short Text Conversion system is evaluated based on its ability to generate context-rich, coherent, and semantically relevant text from short inputs. Experimental results indicate that traditional text generation methods produce generic and less informative outputs due to limited input context. In contrast, the proposed semantic key-controlled approach significantly improves the quality of generated text by incorporating meaningful semantic features. Evaluation metrics such as BLEU score, ROUGE score, and semantic similarity confirm the

effectiveness of the model in generating accurate and relevant content.

A comparative analysis was conducted between different models, including basic Seq2Seq models, attention-based models, and transformer-based architectures. Results show that Seq2Seq models without attention struggle to retain context, leading to incomplete or irrelevant outputs. The introduction of attention mechanisms improves performance by focusing on important parts of the input. Transformer-based models, such as GPT, achieve the highest performance due to their ability to capture long-range dependencies and generate fluent text. The addition of semantic key control further enhances these models by guiding the generation process, ensuring that the output remains aligned with the input intent.

The system was tested with various types of short text inputs, including queries, social media posts, and conversational inputs. Results demonstrate that the system consistently generates meaningful and context-aware outputs across different scenarios. However, challenges such as controlling output diversity, avoiding repetitive content, and handling ambiguous inputs remain. Despite these limitations, the overall system performance is efficient and scalable. The results confirm that integrating semantic key extraction with advanced sequence generation models provides a powerful solution for enriching short text in real-world applications.

V.CONCLUSION

The proposed Rich Short Text Conversion using Semantic Key-Controlled Sequence Generation presents an effective solution for enhancing short and context-limited text into meaningful, coherent, and informative content. By integrating semantic key extraction with advanced sequence generation models such as Seq2Seq and transformer-based architectures, the system overcomes the limitations of traditional text generation methods. The use of semantic keys ensures that the generated output remains relevant and aligned with the original input, improving both contextual understanding and text quality.

Experimental results demonstrate that transformer-based models combined with semantic control mechanisms significantly outperform conventional approaches in terms of accuracy, fluency, and semantic relevance. The system effectively handles various types of short text inputs, including queries and conversational data, making it suitable for applications such as chatbots, content generation, recommendation systems, and intelligent assistants. Additionally, the incorporation of attention mechanisms and feature fusion techniques further enhances the

system's ability to generate high-quality outputs.

In conclusion, the proposed system provides a scalable, efficient, and intelligent approach to short text enrichment. While challenges such as computational complexity, ambiguity in input text, and output diversity remain, the benefits of improved text generation and contextual understanding outweigh these limitations. Future work may focus on optimizing model efficiency, integrating real-time processing, and incorporating more advanced controllable generation techniques. Overall, this study highlights the importance of combining semantic understanding with deep learning models to advance natural language processing applications.

REFERENCES

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proc. ICLR*, 2013.

[2] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in *Proc. EMNLP*, 2014, pp. 1532–1543.

[3] I. Sutskever, O. Vinyals, and Q. Le, "Sequence to Sequence Learning with Neural Networks," in *Proc. NeurIPS*, 2014, pp. 3104–3112.

[4] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *Proc. ICLR*, 2015.

[5] A. Vaswani et al., "Attention Is All You Need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[6] A. Radford et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[9] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, 2010.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[11] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015.

[12] F. Chollet, "Keras," 2015.

[13] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL*, 2019, pp. 4171–4186.

[14] T. Brown et al., "Language Models are Few-Shot Learners," in *Proc. NeurIPS*, 2020, pp. 1877–1901.

[15] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *JMLR*, 2020.

[16] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019.

[17] A. Radford et al., "Improving Language Understanding by Generative Pre-Training," OpenAI, 2018.

[18] J. Brownlee, *Machine Learning Mastery With Python*. Machine Learning Mastery, 2016.

[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[20] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation," in *Proc. IJCAI*, 1995, pp. 1137–1143.