Development of a Korean Counseling Chatbot : Data Augmentation and Stable Training with Transformers

Anonymous Author(s)
Affiliation
Address
email

Abstract

This paper presents the development of a Korean counseling chatbot using Transformer-based models. By augmenting and preprocessing the data, a stable training procedure was devised to mitigate overfitting. The results show that the chatbot provides relevant responses to user queries, including those related to comfort and emotional support.

1 Introduction

Can AI be a comfort to someone?

1.1 Background and Motivation

Mental health issues are becoming increasingly recognized as a critical global concern, and the demand for accessible mental health care is rising. Traditional therapy and counseling are often not easily accessible due to factors such as cost, time, and stigma. With the rapid development of artificial intelligence (AI), counseling chatbots have emerged as promising tools to provide immediate, accessible, and cost-effective support to individuals seeking mental health assistance.

In particular, Korean counseling chatbots face unique challenges due to the linguistic and cultural complexities of the Korean language. Korean is a high-context language, where much of the meaning in conversation relies not only on the literal interpretation of words but also on understanding the emotional tone, context, and subtle nuances. Therefore, creating an AI-powered chatbot that can effectively engage in therapeutic dialogue in Korean is a complex task that requires advanced natural language processing (NLP) techniques.

The margins in 2024 are the same as those in previous years.

1.2 Research Objectives

This research focuses on developing a Korean counseling chatbot that can offer effective emotional support while overcoming the linguistic and cultural barriers inherent in the Korean language. The specific objectives of this study are:

1. Data Augmentation: To enhance the quality and diversity of training data, we explore methods to generate synthetic data, addressing the issue of data scarcity for training deep learning models.
2. Transformer-Based Model: We leverage Transformer models, which have demonstrated state-of-the-art performance in various NLP tasks, to train the chatbot. The goal is to ensure the chatbot can generate natural, contextually relevant, and empathetic responses in Korean.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

3. Stable Training: To ensure that the model can be trained effectively without overfitting, we introduce strategies such as regularization and early stopping, ensuring stable learning on small or imbalanced datasets.

By achieving these objectives, we aim to create a counseling chatbot that is capable of providing high-quality support for mental health in Korean, which could be scaled to a broader audience, offering much-needed services for those unable to access traditional forms of therapy.

2 Related Works in Counseling Chatbots

The development of counseling chatbots has been an area of significant interest in recent years. Several approaches and studies have contributed to advancing the capabilities of these systems.

1. Early Models and Approaches: One of the earliest and most well-known chatbot models for mental health was ELIZA (Weizenbaum, 1966). ELIZA was a rule-based chatbot that mimicked a Rogerian psychotherapist by reflecting user inputs. While ELIZA was groundbreaking at the time, its lack of deep understanding and empathy led to limited effectiveness. Since then, many other chatbot systems have been developed, incorporating more advanced techniques such as natural language processing (NLP) and machine learning (ML).
2. Machine Learning and Deep Learning Approaches: More recent chatbot systems have utilized machine learning techniques to improve the interaction quality. For example, Woebot, a mental health chatbot based on CBT, uses deep learning to understand and respond to emotional cues in user inputs. Woebot has been shown to be effective in helping users manage symptoms of anxiety and depression, primarily by delivering therapeutic interventions through conversation. These systems often rely on sequence-to-sequence models or transformer-based architectures, which enable them to generate contextually appropriate responses.
3. Transformer-Based Models in Counseling Chatbots: The advent of Transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformers), has greatly improved the ability of chatbots to understand and generate natural language. Transformer models have achieved state-of-the-art results in many NLP tasks, including text classification, sentiment analysis, and response generation. These models, with their attention mechanisms, can capture long-range dependencies and nuances in conversation, making them highly effective for building emotionally intelligent chatbots capable of offering counseling support.
4. Challenges in Multilingual Counseling Chatbots: The effectiveness of a counseling chatbot is highly influenced by the language in which it operates. For instance, most of the well-known counseling chatbots have been developed in English-speaking contexts, limiting their accessibility for non-English speakers. Research on multilingual and culturally aware chatbots is still in its early stages. In particular, Korean language counseling chatbots face unique challenges due to the complexities of the Korean language, including its rich honorifics system, nuanced emotional expressions, and cultural norms regarding mental health.
5. Empathy and Sentiment Recognition: Several studies have focused on improving the chatbot's ability to recognize and respond to emotional cues in text. For example, EmoReact, a sentiment analysis model, is trained to detect emotions such as happiness, sadness, and anger in textual input. By combining sentiment analysis with natural language generation (NLG), chatbots can tailor their responses to match the emotional state of the user, helping to create a more empathetic and supportive experience.
6. Data Scarcity and Data Augmentation in Counseling Chatbots: A significant challenge in building effective counseling chatbots is the scarcity of domain-specific data. Mental health conversations are private and sensitive, making it difficult to obtain large datasets for training. Several studies have explored data augmentation techniques to overcome this challenge. For example, back-translation and text synthesis methods can be employed to create synthetic training data, thus enriching the chatbot's understanding of a wide range of user inputs.

Cultural Considerations in Counseling Chatbots An often overlooked but crucial factor in counseling chatbot development is cultural context. Mental health and emotional expression vary significantly across cultures, and understanding these differences is vital for creating a chatbot that resonates with users.

- **In Korea**, there is a strong societal stigma surrounding mental health issues, which can lead to reluctance in seeking professional help. As a result, counseling chatbots offer a low-stigma alternative for individuals who may feel uncomfortable seeking traditional therapy.
- **Language-specific nuances** in Korean, such as formal and informal speech, emotional expression, and indirect communication styles, further complicate the development of effective chatbots. A Korean chatbot must be able to engage users while respecting these nuances and delivering supportive responses that reflect a deep understanding of the Korean cultural context.

3 Model Architecture and Training Methodology

3.1 Transformer

In this study, we implemented a Korean chatbot based on the Transformer architecture. The model adopts an Encoder-Decoder structure, with each component designed as follows:

Encoder: After embedding the input sentence, it extracts contextual information through Multi-Head Self-Attention Mechanisms and Position-wise Feedforward Networks.

Decoder: It receives the encoder's output along with previously generated words to predict the next word. The decoder also consists of Multi-Head Self-Attention Mechanisms and Position-wise Feedforward Networks.

This structure enables the model to effectively understand the meaning of input sentences and generate natural Korean responses.

In addition, the Transformer model was trained directly from the training data, without any pre-training model, and achieved stable convergence and overfitting through proper hyperparameter tuning.

3.2 Data preprocessing and augmentation

In this study, we used the publicly available Korean dialog dataset for chatbot training.

https://github.com/songys/Chatbot_data/blob/master/ChatbotData.csv

This dataset is widely used in Korean language chatbot research for its simplicity and relevance to daily conversation.

Overview: The dataset consists of **11,876 pairs of user questions and chatbot responses**. Each data point includes:

- Q: User question (input sentence)
- A: Chatbot response (target sentence)
- label: Category label (0, 1, or 2) representing the type of question

Characteristics:

- Contains short, colloquial Korean sentences suitable for training conversational agents.
- The label indicates the conversational intent, but our primary focus was on Q–A pairs for generative chatbot modeling.
- Openly available under a permissive license and commonly used as a baseline dataset.

Data Preprocessing :

- Special Character Removal: Unnecessary special characters were removed from the text to improve data quality.
- Tokenization : We used SentencePiece to tokenize sentences into subword units, alleviating the rare word problem and improving the model's ability to handle various words. SentencePiece was trained using the Byte-Pair Encoding (BPE) algorithm.

Data Augmentation :

- Lexical Substitution : Lexical Substitution involves replacing words in a sentence with semantically similar alternatives. Using word embedding similarity, we analyzed each input sentence and substituted specific words with their nearest neighbors in the embedding space.

4 Experimental Results and Performance Evaluation

In this experiment, we trained and evaluated a Korean counseling chatbot based on the Transformer model. We applied lexical substitution as a data augmentation technique to improve the model's generalization ability.

4.1 Experiment Setup

The dataset used in the experiment is based on Songys' Chatbotdata, which consists of question-answer pairs. During the data preprocessing phase, the SentencePiece tokenizer was used to tokenize the text, and the Transformer model with embedding and encoder-decoder structure was trained. The training was conducted in a GPU environment, and hyperparameters such as learning rate, batch size, and epoch were optimized for the best performance.

4.2 Metrics

The performance of the model was evaluated mainly using accuracy and BLEU score. Accuracy measures how well the chatbot provides appropriate responses to the given inputs, while the BLEU score is commonly used in machine translation tasks to evaluate the quality of generated text.

4.3 Training Results

The training lasted for 20 epochs, and the model was evaluated using a validation set. The results showed that the Transformer model gradually improved its performance from the early stages of training, achieving an accuracy of 87 percents and a BLEU score of 0.44. The model was trained with methods such as early stopping to prevent overfitting and ensure stable learning.

4.4 Model Performance

The final trained model was able to generate natural and appropriate responses to actual user queries. For example, when the user asked "I am Boring", the chatbot generated an appropriate response like "Is it frustrating" This shows that the model had learned adequately, and the data augmentation technique contributed to improving its generalization performance.

4.5 Result Analysis

The experiment confirmed that the Transformer model performed well in text generation and response creation. Moreover, applying data augmentation helped the model handle various contexts and generate more accurate responses. However, there were cases where the chatbot's responses were not perfect, indicating that further improvements in data augmentation techniques and more epochs are necessary in future studies.

5 Insights and Future Directions

The Korean consultation chatbot developed in this study is based on the Transformer model, and we found that it is able to provide natural and empathetic responses in line with the user's emotions.

Through data augmentation techniques, the model learned a variety of expressions, which allowed it to generate consistent responses to new inputs. Nevertheless, there are some limitations and future improvements.The Korean consultation chatbot developed in this study is based on the Transformer model, and we found that it is able to provide natural and empathetic responses in line with the user's emotions. Through data augmentation techniques, the model learned a variety of expressions, which allowed it to generate consistent responses to new inputs. Nevertheless, there are some limitations and future improvements.

5.1 Future research directions

This study used a basic Transformer model for a counseling chatbot and can suggest various research directions for future improvements and extensions.

- **Expand and diversify dataset**: We can increase the versatility and reliability of model by building a conversation dataset that includes a more diverse user base. In particular, we can improve model's ability to respond more accurately and flexibly by adding data that reflects different emotional states and situations.
- **Handling multi-turn conversations**: In real-world counseling, multiple conversations take place, requiring skills to understand the context of multi-turn conversations and generate appropriate responses. To this end, it is important to develop model architectures and training methods that can reflect the flow of multi-turn conversations in current models. For example, we can extend Transformer's Encoder-Decoder structure to track the flow of a conversation and improve it to generate responses that reflect previous conversations.
- **Personalized feedback system**: To provide more personalized advice tailored to the user's situation and emotions, research is needed to generate customized responses that reflect the user's previous conversations and emotional state. For example, if a user repeatedly expresses a certain emotional state, we can ensure that the model recognizes this and provides personalized advice or feedback accordingly.
- **Improve efficiency**: Current models require large datasets and computational resources. Therefore, to apply chatbots to real-time consultation systems, we need to improve the lightweight and efficiency of the models. We need to study how to maintain high performance with smaller models and explore how they can operate smoothly in mobile or resource-constrained environments.

5.2 Concolusions

In this study, we developed a Korean consultation chatbot using the Transformer model, and improved the model performance through data augmentation techniques and stable learning methodology. However, the current model has some limitations, and future research can develop a more improved consultation chatbot through data augmentation, multi-turn conversation processing, and personalized response system. If these studies are carried out, it will be possible to provide more helpful chatbots for people in real consultation systems.

References

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008. *In this foundational paper, the Transformer architecture is introduced, which has become the core model for many natural language processing tasks. The self-attention mechanism and positional encodings proposed in this paper address the shortcomings of previous sequence-based models like RNNs.*

[2] TensorFlow, "Transformer Tutorial," *TensorFlow Documentation*, 2020. [Online]. Available: https://www.tensorflow.org/text/tutorials/transformer?hl=ko *The TensorFlow Transformer tutorial provides a detailed implementation guide for building transformer models for sequence-to-sequence tasks. This resource includes code examples for tokenization, embedding layers, multi-head attention, and model training, making it a helpful starting point for anyone working with transformer-based architectures.*

[3] Google, "Transformer Model in Trax," *Trax GitHub Repository*, 2020. [Online]. Available: https://github.com/google/trax/blob/master/trax/models/transformer.py *The Trax repository by Google offers an efficient implementation of the Transformer model, focusing on both performance and scalability. It includes various optimization techniques that enhance the model's ability to handle large-scale datasets, making it suitable for research and production applications.*

[4] Kyubyong, L., "Word Vectors," *GitHub Repository*, 2020. [Online]. Available: https://github.com/Kyubyong/wordvectors *This GitHub repository provides various pre-trained word vectors and tools for using them in natural language processing tasks. The repository includes vectors trained on large corpora, such as Wikipedia, and allows easy integration into custom NLP projects.*

[5] Jang, H., & Lee, S. H., "A study on the development of a Korean counseling chatbot using deep learning," *Journal of Korean Society for Computer Application*, vol. 19, no. 4, pp. 12-20, 2019. DOI: 10.5392/JKCA.2019.19.04.012. *This paper discusses the development of a Korean counseling chatbot using deep learning techniques. The authors explore the use of NLP models for understanding user input and delivering meaningful responses in a counseling context, which is particularly challenging in Korean due to language and cultural nuances.*