Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766327

DESIGNING SBERT AND ENSEMBLE LEARNING MIXTURE MODEL FOR SENTIMENT ANALYSIS OF AUTISTIC CHILDREN

Hyeonji Kim [1], Yoosoo Oh [2, 1]

[1]Dept. of Information and Communication Engineering, Daegu University, Gyeongsan-si, Republic of Korea

[2]School of AI, Daegu University, Gyeongsan-si, Republic of Korea

Abstract: Autistic children communicate through speech and actions. Children with autism express their emotions primarily through words rather than actions. Therefore, this paper proposes a mixing model that combines Sentence-BERT as a language embedding model and Voting Classifier as a machine learning ensemble model. The proposed mixture model can improve the performance of sentiment analysis models for autistic children by not only sentence embedding but also by giving them semantic weights. This paper proposes a model for the emotional analysis of children with autism in the future.

Keywords; Autistic children, NLP, SBERT, Machine Learning, Ensemble Learning

1 1. Introduction

Developmental disabilities have delays and abnormalities in social relationships, communication, and cognitive development. Autistic children, among the types of developmental disabilities, are characterized by delayed language and communication development. Children with autism express emotions through language and repetitive behavior. The use of immediate or delayed echo words characterizes them. Therefore, it is difficult for carers to grasp the emotions of autistic children.[1]

This research paper uses a language model and a machine learning classification method to classify emotions using language. Among the language models, Sentence-BERT (SBERT) is a model that improves sentence embedding performance by adding pooling operations to the output of BERT. SBERT outputs vectors to all input sentences and derives sentence embeddings with meaning through cosine similarity calculations [3]. Among the machine learning models, the Voting Classifier ensemble model proceeds with learning by placing two or more machine learning models. Finally, the Voting Classifier votes on the machine learning results used for learning to derive the final prediction [9].

In this research work, we propose an S-BEL(Sentence BERT – Ensemble Learning) Mixture model that combines the Voting Classifier model among SBERT and ensemble

[1] Corresponding author: yoosoo.oh@daegu.ac.kr
Received:; Accepted:; Published:

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A07091326)

models. The proposed S-BEL Mixture model derives fixed vectors for all input sentences and gives semantic weights. This paper builds a model using non-disabled people's emotional dialog data because of difficulties in collecting data for children with autism. The proposed S-BEL Mixture model enables the emotional classification of non-disabled people. Furthermore, the proposed S-BEL Mixture model can generate answers in the form of appropriate chatbots corresponding to the analyzed emotions.

2. Related Research

Park Sang-min, Lee Jae-Yoon, Sung Yuri, and Kim Jae-Eun proposed an SBERT model learning method that can express semantic vectors considering the amount of keyword information. Park Sang-min, Lee Jae-Yoon, Sung Yuri, and Kim Jae-Eun constructed {positive and negative} keyword training data by sentence n-gram to construct training data containing keyword information. Park Sang-min, Lee Jae-Yoon, Sung Yuri, and Kim Jae-Eun used the SBERT model to learn the data from the collected learning sentences and constructed keyword pairs. In this related study, an SBERT model trained on data reflecting keywords showed a 2.74% improvement in performance compared to the existing SBERT model. [2]

Yoon Hye-jin, Ku Ja-hwan, and Kim Eung-mo used five types of word embedding techniques and three types of machine learning classification models to advance sentiment classification. They used CounterVectorizer, TfidfVectorizer, Word2vec (CBOW), Word2vec (Skip-gram), and Pretrain_Word2vec for five-word embeddings, and Decision Tree, and also used RandomForest, and Logistic Regression models as machine learning classification models. As a result of testing, related research showed that CounterVectorizer outperforms TfidifVectorizer. Yoon Hye-jin, Koo Ja-hwan, and Kim Eung-mo conducted an accuracy comparison according to the number of emotion categories. As a result, they confirmed that the accuracy improved by more than 30% when two emotion categories (positive and negative) were used compared to when seven emotion categories were used. In addition, they confirmed that a subdivided emotion category was unsuitable for emotion classification through embedding. [4].

Park Sang-min, Lee Jae-yoon, Son Yu-ri, and Kim Jae-eun made two keywords through ngram and paired keywords and sentences to learn. He then analyzed emotions by applying three types of machine learning: Yoon Hye-jin, Koo Ja-hwan, and Kim Eung-mo. In addition, they also confirmed the improvement of the analysis model accuracy according to the number of emotion categories. People's emotions can be expressed in various ways. Some emotions are inappropriate to be divided into positive and negative categories. Since emotion models are specific and diverse, better performance can be obtained by combining each machine learning rather than using a single model to improve the accuracy of emotion machine learning. Therefore, in this paper, the keywords are subdivided into four emotions to proceed with training. In addition, this paper proposes a model combining several machine learning models using natural language processing models and ensemble techniques.

3. S-BEL Mixture Model

This paper constructed a model using more than 20,000 emotional datasets provided by the National Information Society Agency. The National Information Society Agency's emotional data has six major categories ("joy," "embarrassment," "anger," "anxiety," "wound," and "sadness") emotions, human response, and system response. In this paper, we use human response one and system response 1. In addition, this paper includes the wound category as the sadness category and the panic category as the anger category. Therefore, we use a total of four emotion categories ("joy," "anxiety," "anger," and "sadness"). The S-BEL Mixture model diagram of the model proposed in this paper is shown in Figure 1 below. The proposed model consists of a Sentence Bert model, an ensemble machine learning model, and an S-BEL Mixture model.

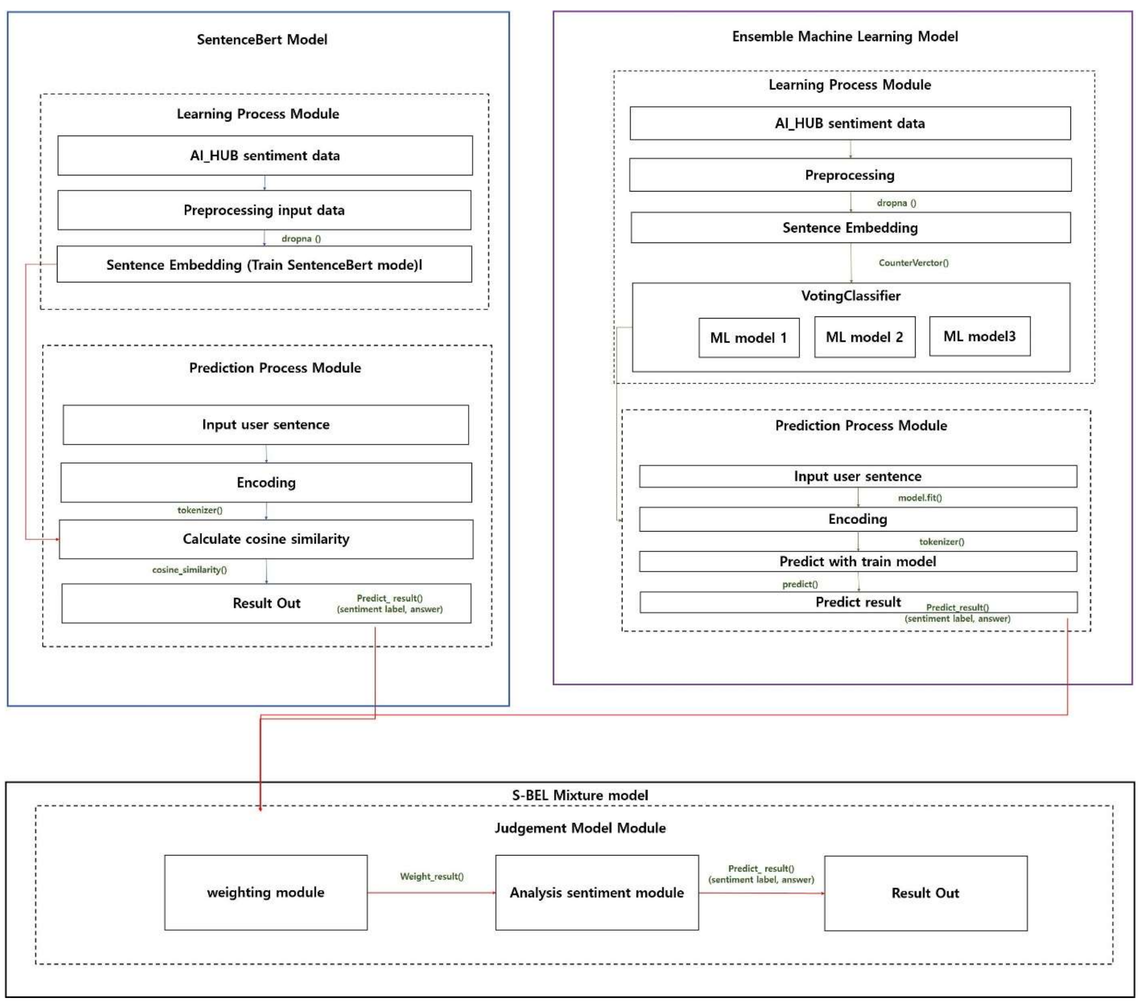

Figure 1. S-BEL Mixture Model Diagram

(ML1: MultinomialNB, ML2 : RandomForest, ML3 : XGBoost)

The S-BEL hybrid model proposed in this paper uses the predictive results of SBERT models and ensemble machine learning models. These paperweights the predicted result values through the S-BEL Mixture model. The SBERT model proceeds with training through the STS dataset, which measures sentence similarity, and the NLI dataset, which identifies the relationship between sentences. In this paper, we learn the fine-tuning Sentence Bert model using the data consisting of human sentence one and emotion large category pairs as feature values. The trained SentenceBERT proceeds with an encoding vectorizing the sentences the user enters to make predictions corresponding to user input. Sentences vectorized by encoding predict the closest sentence and emotion by cosine similarity calculation with the trained model. Cosine similarity is calculated as Equation 1 [8].

$$\text{similarity} = \cos(\theta) = \frac{A * B}{\| A \| \| B \|}$$

Equation 1. Cosine Similarity Formula

Ensemble learning uses two or more machine learning models among machine learning classifications. The ensemble learning model combines the prediction results of the models used to derive one final prediction result. The VotingClassifier models constructed in this paper use XGBoost, RandomForest, and MultinomialNB models. A Hard-Voting model finalizes the predicted value of the class that gets the most votes among the results of each classifier model by majority classification. In contrast, a Soft-Voting model predicts the class with the highest average probability for each classifier prediction exists.[7] For the proposed model, we applied the Soft-Voting model. In this paper, we make sentences vectorized by the number of occurrences of words through CounterVectorizer for collected emotional data. Afterward, we trained the MultinomialNB, RandomForest, and XGBoost models to the VotingClassifier model through the pipeline. The proposed ensemble model proceeds with embedding user input sentences to predict emotions corresponding to user input. The emotional result of the input sentence is output through the model's prediction according to the probability.

The weighting formula for the S-BEL Mixture model proposed in this paper is shown in Equation 2 below. In the proposed model, b_s is the similarity measure of the SBERT model, W_ms is the accuracy value of the ensemble learning model, and X_mp is the probability value measured by the ensemble learning model. Equation 2 performed a correlation analysis of accuracy and prediction probability via the product of W_ms and X_mp. In addition, equation 2 shows the deflection value via the value of b_s−1. Furthermore, Equation 3 applied a sigmoid function formula for scaling between 0 and 1 [10].

$$S.BEL_w = W_{ms} * X_{mp} + (b_s - 1)$$

Equation 2. S-BEL Mixture Model Weight Formula

($S.BEL_w$: Weight result value,

W_{ms}: $ELaccuracyvalueofthemodel$, X_{mp} : Probability value measured through EL model, b_s: $SimilaritymeasureofSBERTmodel$)

$$S.BEL_w = \frac{1}{e^{-(S.BE\ \ w)}}$$

Equation 3. Weight scaling formula of S-BEL Mixture model

The proposed S-BEL Mixture model constructed a decision module using a threshold. This paper used an accuracy comparison by designating a threshold from 0.1 to 1.0, finding an appropriate threshold. Figure 2 shows that the threshold value of 0.5 derived the highest accuracy. Therefore, the proposed S-BEL mixture model verifies whether the S-BEL mixture weight value is above or below the specified threshold (0.5) through the judgment module. If the decision result of the decision module is greater than or equal to the specified threshold, an emotion label and chatbot-type answer corresponding to the emotion predicted by the proposed model are output. Conversely, when the determination result of the decision module is less than the specified threshold value, the result value is not output.

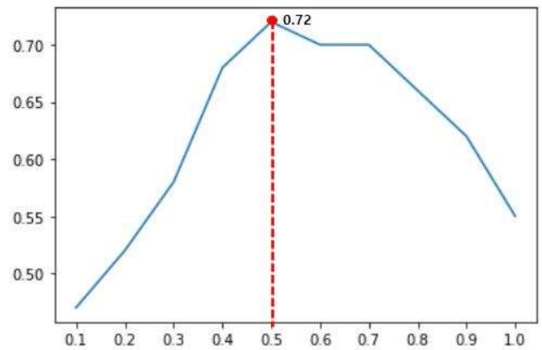

Figure 2. Performance evaluation results according to the threshold

2 4. Experiments and Results

This paper compares the accuracy of the classification model and the proposed model, as shown in Figure 3. In this paper, we measured accuracy by dividing the train data and the test dataset in a 7:3 ratio and dividing the total number of samples by the correctly predicted number. As a result of the experiment, the accuracy was higher when progressing using ensemble learning than when progressing with sentiment analysis using each machine learning algorithm. Furthermore, the results ensured that the proposed S-BEL Mixture model achieved the highest accuracy by comparing the compared models. We obtained an accuracy of about 93% in the sentiment analysis performed.

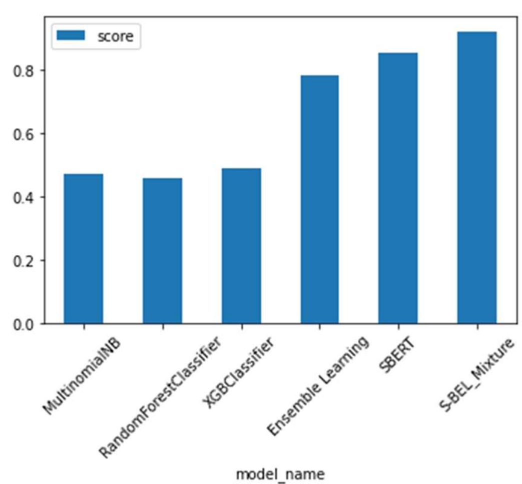

Figure 3. Benchmark Model Accuracy Comparison

For experimental evaluation, we utilized BLEU and Cosine-Similarity evaluations to ensure the accuracy of the chatbot-type response of the S-BEL Mixture system. BLEU (Bilingual Evaluation Understudy) is one of the methods for evaluating the performance of machine translation systems. [11] BLEU measures performance by comparing how similar the result of the machine translation system is to the actual result. The BLEU evaluation method means that the higher the score, the better the performance.[11] Cosine-Similarity measures the similarity between the generated sentence and the correct answer sentence. Cosine-Simlarity is measured through the angle of two vectors and ranges from -1 to +1. The closer Cosine-Similarity is to +1, the highest the similarity.[12]

To verify the evaluation method proposed, we conducted a comparative analysis of the BLEU Score and Cosine Similarity Score for each data. Figure 4,5,6,7,8,9 shows each algorithm's BLEU and Cosine Similarity Score scores according to the number of test data sets. Figure 4 is BLEU Score and Cosine Similarity Score Graph about S-BEL Mixture Model. Figure 5 is the BLEU Score and Cosine Similarity Score Graph about the Ensemble Model consisting of MultinomialNB, RandomForesClassifier, and XGBoost. Figure 6 is the BLEU Score and Cosine Similarity Score Graph about the Ensemble Model consisting of MultinomialNB and RandomForestClassifier. Figure 7 is the BLEU Score and Cosine Similarity Score Graph about the Ensemble Model consisting of MultinomialNB, XGBoost. Figure 8 is the BLEU Score and Cosine Similarity Score Graph about the Ensemble Model consisting of RandomForesClassifier, XGBoost. Figure 9 shows the correlation between the BLEU score of the S-BEL Mixture Model and the cosine similarity score. The heat map on the right side of Figure 9 shows the correlation between S-BEL Mixture Model's BLEU score and cosine similarity score using the Pearson correlation coefficient. The Pearson correlation coefficient represents a linear relationship between two variables. Pearson correlation coefficients indicate that as the correlation coefficient value approaches 1, the relationship between the two variables is higher.[13]

Figures 4, 5, 6, 7, 8, and 9 show that the evaluation score differs depending on the number of data sets. Since the text of the correct answer heavily influences BLEU and Cosine Similarity Scores, the accuracy decreases when an answer is generated by using a word that is not in the correct answer in the test data set. Figure 10 shows the correlation between the proposed S-BEL Mixture model's BLEU Score and the Cosine Similarity Score. The heat map graph on the right shows that the BLEU Score and the Cosine Similarity Score have a high correlation of 0.88. The heat map graph means that the BLEU Score and the Cosine Similarity Score derive scores in a similar way. Therefore, the verification method proposed in this paper has a significant meaning.

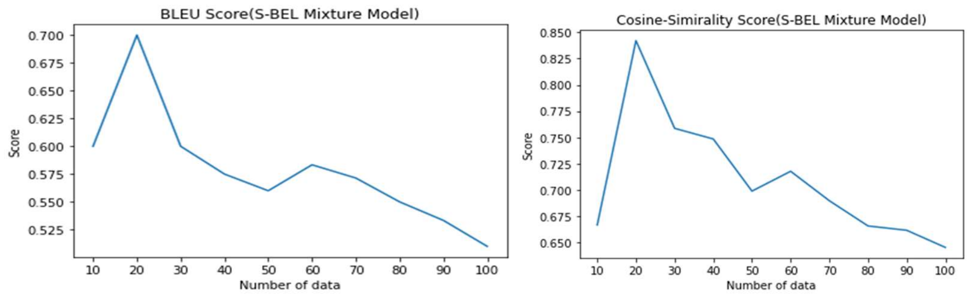

Figure 4. S-BEL Mixture Model BLEU, Cosine-Similarity Score

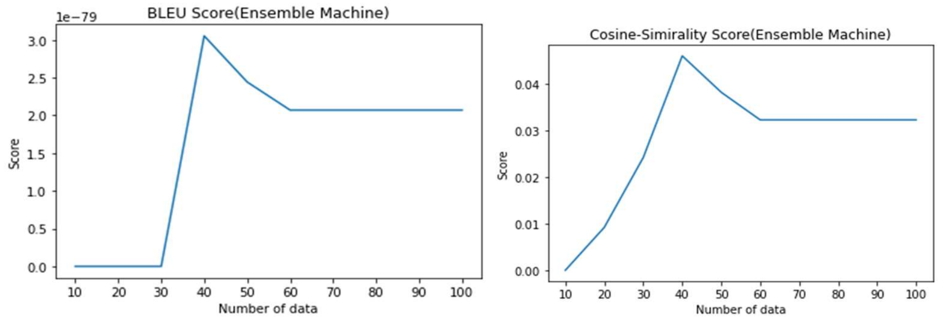

Figure 5. Ensemble Machine (MultinomialNB, RandomForesClassifier, XGBoost) BLEU, Cosine-Similarity Score

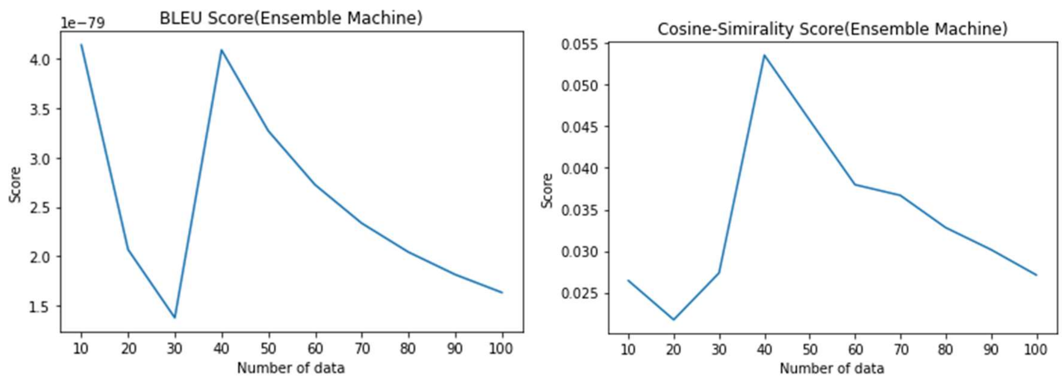

Figure 6. Ensemble Machine (MultinomialNB, RandomForesClassifier) BLEU, Cosine-Similarity Score

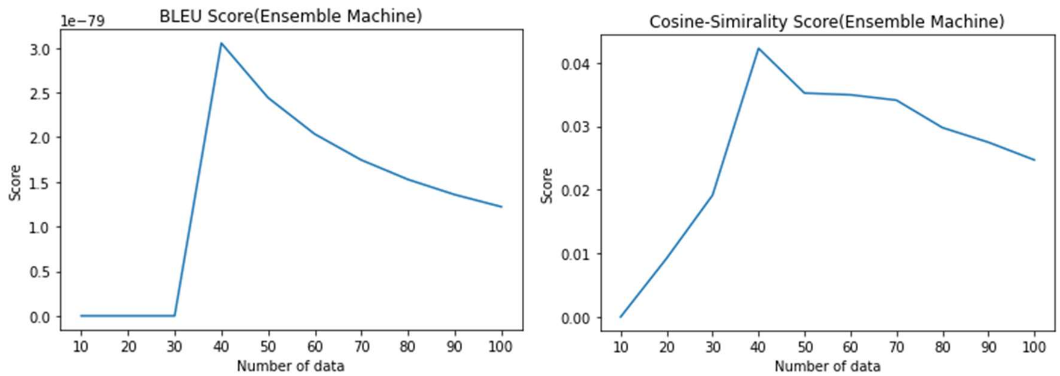

Figure 7. Ensemble Machine (MultinomialNB, XGBoost) BLEU, Cosine-Similarity Score

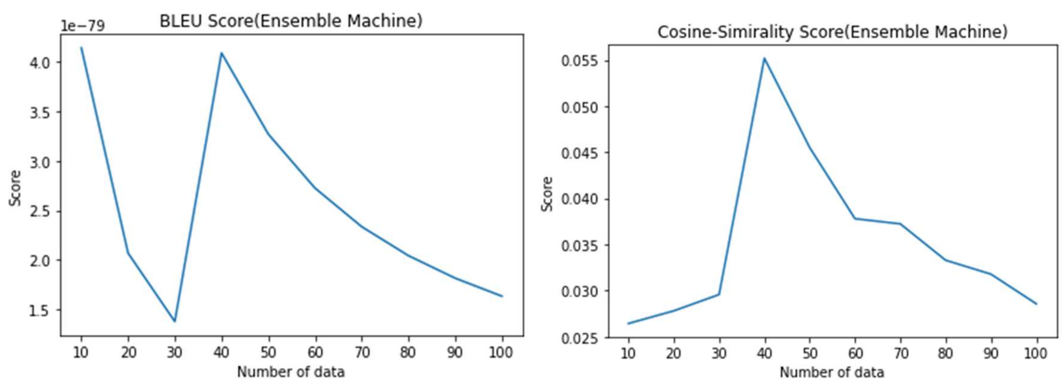

Figure 8. Ensemble Machine (RandomForesClassifier, XGBoost) BLEU, Cosine-Similarity Score

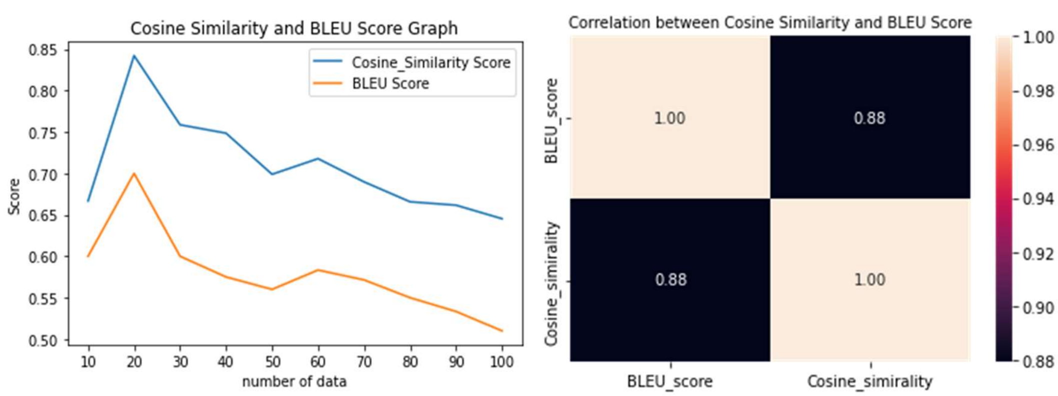

Figure 9. Correlation between Cosine Similarity and BLEU Score Graph

In this research paper, we evaluated a dataset of 60 emotion sentence tests with the highest score through the evaluation verification method. As a result of the experiment, as shown in Table 1, it can be confirmed that the proposed S-BEL Mixture model derives high performance in all performance evaluations with a BLEU value of 0.593 and a Cosine-Similarity of 0.712.

Table 1. Measurement of the accuracy of each system's chatbot-type response

System name	BLEU Score	Cosine-Similarity Score
S-BEL Mixture system	0.593	0.712
Ensemble Machine (MultinomialNB, RandomForesClassifier, xgboost)	0.020	0.032 (1e-79)
Ensemble Machine (MultinomialNB, RandomForesClassifier)	0.016	0.027 (1e-79)
Ensemble Machine (MultinomialNB, xgboost)	0.011	0.024 (1e-79)
Ensemble Machine (RandomForesClassifier, xgboost)	0.016	0.028 (1e-79)

Table 2 below is a sample result of the proposed S-BEL mixture model system. Through the resulting sample, this paper confirmed whether the actual emotion label for the user input sentence matches the emotion label predicted by the proposed mixed model.

	Input Sentence	Real sentiment labels	Predict sentiment labels	Answers in the form of chatbots
1	Sad	Sadness	Sadness	You must be sad.
2	That's very annoying.	Anger	Anger	You're annoying.
3	I'm angry.	Anger	Anger	I think you'll be so angry.
4	Glad.	Joy	Joy	Oh, you must be so happy!
5.	Are you living a happy married life?	Joy	JOY	After getting married, my life has become so comfortable

Table 2. Experience Result (Excerpts of some sentences from the test data)

Table 2 shows that the actual and predicted emotion labels resulting from the input sentences are almost identical. In addition, the proposed model confirms that the actual emotion labels of simple sentences such as 'I am angry' and 'I am happy' match the predicted emotion labels. Therefore, the proposed model is capable of appropriate chatbot-style response output and emotion label prediction.

3 5. Results and Discussion

In this research paper, we proposed an emotion analysis model using SBERT, ensemble machine learning, and S-BEL Mixture weights. Our proposed S-BEL Mixture model had the highest accuracy compared to other emotion analysis models for analyzing emotion. The currently proposed sentiment analysis model used the sentiment data of non-disabled people rather than the dialogue data of autistic children for sentiment analysis. For future research, we plan to collect conversation data about children with autism and train using the data collected. Then we will fine-tune the SBERT according to the training data. Therefore, future research will improve the performance of learning models by optimal fine-tuning. In addition to models such as RandomForest and XGBoost used to improve the accuracy of the VotingClassifier model, we plan to combine various machine learning models to present the optimal combination and proceed with learning. The correct answer data set greatly influences the verification method used in this paper. Therefore, we will verify this in future studies by pre-processing the correct answer data set. We also plan to extend the proposed mixture model to improve its performance of the model.

4 References

[1] JeongMi Lee. "Study on Spontaneous Communicative Behavior in Children with Autism." Emotional and Behavioral Disorder Research 23, no.4 (2007): 345-359.

[2] SangMin Park, JaeYoon LeeYu Ri Sung, JaeEun Kim. (2022). SBERT-IQ: A Sentence-BERT-based embedding model considering keyword information content. Proceedings of the Korean Information Science Society, (), 1058-1060.

[3] Nils Reimers, Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," In Proceedings of Empirical Methods in Natural Language Processing, pp. 3980-3990, 2019.

[4] HyeJin Yoon, JahWan Koo, and UngMo Kim. "Comparison of Word Embedding Performance for Sentiment Classification." Proceedings of the Annual Meeting of the Korean Information Processing Society 28.2 (2021): 760-763.

[5] Y. J. Jang, H. J. Kim, and H. J. Jo, "Data Mining," KNOU PRESS, 2016, 1-200

[6] W. Yun, D.H. Seo, S. Min, H. Nam, "Breast Cancer Classification using RandomForest and XGBoost," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 113-114, 2021.

[7] M. C. Jeong, J. Lee, and H. Oh, "Ensemble Machine Learning Model Based YouTube Spam Comment Detection," Journal of the Korea Institute of Information and Communication Engineering, vol. 24, no. 5, pp. 576–583, May 2020.

[8] SungEun Gil. (2020). Research on action-based authentication by cosine similarity measure Proceedings of the Annual Meeting of the Korean Information Processing Society, 27(1), 165-168.

[9] HuiWon Song, JuHyun Lee, SungHae Jun, SangSung Park. (2022). A Study on Inferable Patent Evaluation Model using Ensemble Method. Journal of Korean Institute of Intelligent Systems, 32(5), 379-385.

[10] Sanjeev Kumar, Mahesh Chandra). "Detection of Microcalcification Using the Wavelet-Based Adaptive Sigmoid Function and Neural Network." Journal of information processing systems 13.4 (2017): 703-715.

[11] Junghyun Cho, Hyunki Jung, Chanyoung Park, and Yuseop Kim. " An Autonomous Assessment of a Short Essay Answer by Using the BLEU " Korean HCI Conference 2009.2 (2009): 606-610.

[12] Eungju Kwon, Jongwoo Kim, Nojeong Heo, and Sanggil Kang. "Personalized Recommendation System using Level of Cosine Similarity of Emotion Word from Social Network" Information Technology Architecture Research 9.3 (2012): 333-344.

[13] Dongho Lee, Minseo Lee, Jaeseo Lee, Heejo Park, Hyunsung Jun, Seehwan Yoo.(2021).Real-time adaptive model training technique using Pearson correlation coefficient. Proceedings of the Korean Information Science Society Conference,1536-1538.