Customer Response Analysis Based on NLP of Sales Activity Texts

Hee Choi
Dept. of AI Techno Convergence
Soongsil University
Seoul, Korea
jameschoi@soongsil.ac.kr

Jeong-sook Ha
AI Convergence Research Institute
Soongsil University
Seoul, Korea
jsha1026@ssu.ac.kr

Kwang-young Park
Dept. of AI Techno Convergence
Soongsil University
Seoul, Korea
1004pky@ssu.ac.kr

Gye-young Kim
Dept. of AI Techno Convergence
Soongsil University
Seoul, Korea
gkkim11@ssu.ac.kr

***Abstract*—This study aims to develop an automatic customer response classification system using NLP (Natural Language Processing) techniques in a B2B sales environment. By utilizing unstructured text data such as meeting notes, emails, and chat records, the study aims to optimize sales strategies through accurate analysis of customer response. The proposed model is a deep learning-based architecture using BERT (Bidirectional Encoder Representations from Transformers) embeddings and an Attention mechanism within a BiLSTM (Bidirectional Long Short-Term Memory) model, classifying customer responses into positive, neutral, and negative categories. The model was trained on a total of 2,865 augmented text samples and achieved a validation accuracy of 89.35%, showing particularly high performance in classifying negative responses. Performance metrics such as Accuracy, Precision, Recall, and F1 Score were used, and the proposed model demonstrated overall superiority compared to existing models. Despite the limitations of domain specificity and response classification categories, the proposed approach shows significant potential for enhancing B2B sales decision-making. Future research should focus on expanding data diversity and integrating multi-modal data to enhance customer response analysis.**

***Keywords*—B2B Sales, Sales Activity, NLP, Text Classification**

I. Introduction

The importance of Sales CRM (Customer Relationship Management) systems is increasingly emphasized across various industries. As the demand for CRM systems continues to grow, companies must efficiently manage and analyze customer data to strengthen customer relationships and improve sales performance [1]. Unstructured text data generated from sales activities, such as meeting records, emails, and chat logs, are valuable assets [1]. In B2B sales, these records capture interactions with customers and play a crucial role in understanding customer needs and response. This data is essential for identifying customer needs, tracking communication, and optimizing sales strategies.

Analyzing customer responses has become essential for developing future sales strategies [2]. However, previous studies mainly focused on customer response analysis or simple classification tasks using structured datasets, which limited their applicability in real-world sales environments. With advancements in NLP (Natural Language Processing) techniques, it has become possible to effectively analyze large-scale text data and optimize sales activities [3]. Notably, Transformer models exhibit high accuracy in classifying customer responses, making them particularly effective in B2B sales settings [4].

This study aims to develop a system that automatically classifies customer responses into positive, negative, and neutral categories using NLP techniques for B2B sales activity text data. This system is designed to provide objective and consistent criteria for classifying customer responses, reducing reliance on subjective judgments by sales representatives.

The study also seeks to optimize sales processes based on customer response data, improving sales strategies, identifying opportunities, and increasing success rates. The ultimate goal is to establish flexible sales strategies that can adapt to changing customer needs in real-time. Additionally, the proposed system aims to verify its effectiveness in real-world B2B sales environments and present practical measures to improve sales performance.

This study focuses on analyzing unstructured text data generated from B2B sales activities using NLP techniques to develop a system for automatically classifying customer responses. The main target data include meeting records, emails, and chat logs, and the goal is to classify customer responses into positive, negative, and neutral categories. The study applies the latest NLP models capable of handling complex text structures and includes comprehensive technical procedures from data preprocessing to model training and evaluation. Through this approach, a practical system applicable to real-world B2B sales environments is developed, and strategies to enhance sales performance are proposed [5].

This study proposes an NLP-based solution applicable to real-world sales environments, aiming to optimize B2B sales processes and improve customer interactions.

II. Related Work

A methodology was proposed to apply customer response analysis to sales activity records generated in B2B CRM systems to predict sales. The primary goal of the study was to utilize CRM data to predict sales performance and analyze the impact of customer response information on sales strategies [6].

The study used actual B2B CRM data to conduct experiments, analyzing the customer response information recorded in CRM sales notes to predict sales performance. It was found that customer response information contained in sales activity notes plays a crucial role in predicting sales outcomes. The proposed methodology analyzed CRM data to identify key attributes influencing deal closure and provided decision support tools to sales representatives. In this process, the Random Forest (RF) model demonstrated the highest performance, and including customer response-related features significantly improved prediction accuracy.

The experimental results showed that the RF model performed well in most experiments, and in particular, the model including customer response-related features greatly improved accuracy and predictive power. The conclusion of the study emphasized that customer response analysis in CRM data can be effectively applied to establishing sales strategies and predicting sales performance and suggested the possibility of further expanding customer response analysis techniques in CRM systems and utilizing them in practice in the future.

In a study related to text classification, a model that combines customer response and semantic information was proposed for customer response classification of user-generated text, and the goal was to perform more precise customer response analysis by utilizing the latest embedding technology (BERT) [7]. SCA-HDNN model consists of three main stages [8]. The first is the text preprocessing stage, where unnecessary elements are removed, and part-of-speech (POS) tagging is performed to enhance the syntactic meaning of the text. The second is the stage of integrating customer response and semantic knowledge, where contextual meaning and customer response information are learned by utilizing BERT (Bidirectional Encoder Representations from Transformers) embedding. The third step is to perform customer response classification of text through the CNN layer, learn the input embedding through the CNN layer and Max-pooling layer, and finally classify the customer response in the Fully Connected layer. The performance evaluation results showed that this model showed superior performance compared to existing customer response analysis techniques. In particular, the fact that context information was reflected more precisely through BERT embedding played an important role in improving performance. In the experiment, the model showed high accuracy and F1 score, proving that integrating customer response and semantic information is very effective for customer response analysis of user-generated text. The conclusion of the study emphasized that the approach of combining semantic knowledge and customer response knowledge is useful for more precisely analyzing the customer response of user text and can be applied to various text analysis applications in the future.

III. Methodology

A. Data Collection

In this study, customer response classification was conducted using text data generated from sales activities. A total of 2,000 text samples were collected, recording meeting details with customers in a B2B sales environment. Among these, irrelevant or system-generated data were excluded from analysis, resulting in 1,574 meaningful data samples.

The collected data were labeled into five categories: 'Highly Positive,' 'Positive,' 'Neutral,' 'Negative,' and 'Highly Negative.' However, due to the small number of 'Highly Positive' and 'Highly Negative' samples, these categories were consolidated into three: 'Positive' (124 samples), 'Neutral' (955 samples), and 'Negative' (495 samples). To balance the dataset, data augmentation was performed to increase the number of samples in each category to 955, resulting in a total of 2,865 samples.

B. Data Preprocessing

The collected text data underwent several preprocessing steps to convert it into an analysis-ready form. First, only Korean text was extracted, and unnecessary sentences or symbols were removed. Additionally, general stop words that were not relevant for customer response analysis were eliminated, and sensitive personal information, such as the names of sales representatives, was removed to maintain data confidentiality.

To analyze text data more effectively, the KoNLPy library's Okt morphological analyzer was used to perform morphological analysis on the text. This enabled the extraction of meaningful lexical information, which was then used to extract features required for customer response classification. Finally, label encoding was performed to convert each text's customer response category into numerical form. The preprocessed data were then split into training (80%) and validation sets (20%) for model training.

C. Model Architecture and Optimization

The proposed model is a deep neural network designed to perform customer response classification based on text data. The model consists of seven primary layers, each responsible for learning features and ultimately classifying customer responses.

The Feature Engineering Layer performs tokenization and padding on text data. This converts each sentence into a fixed-length input, ensuring that the text data is preprocessed in a format suitable for the model.

The Embedding Layer uses BERT embeddings to convert text data into vectors. BERT provides contextually meaningful embeddings, making it a critical component for customer response analysis as it allows for more accurate representation of each sentence's semantic information.

The BiLSTM Layer learns bidirectional contextual information from the text. Initially, a BiLSTM with 256 units is applied, followed by another BiLSTM with 128 units. This layer enables the model to learn both forward and backward contexts of the text, preventing important contextual information from being missed. Dropout probability is set to 0.5 to prevent overfitting by randomly deactivating some units during training.

The CNN Layer learns local patterns in the text using convolution operations. Although CNNs are typically used for image processing, they are also effective for extracting features in text. The CNN layer learns important customer response-related patterns and uses a max-pooling layer to reduce the feature space and efficiently extract key information.

The Attention Layer emphasizes important words and contexts within the text. Through the attention mechanism, weights are assigned to different parts of the text, allowing the model to focus more on crucial information. This learning process enhances the accuracy of customer response classification by capturing the main context of sentences.

The Fully Connected Layer combines the features extracted in the previous layers for final classification. This layer uses the ReLU (Rectified Linear Unit) activation function to add nonlinearity and handle the complexity of customer response classification.

The Output Layer uses the Softmax activation function to classify customer responses into 'Positive,' 'Neutral,' and 'Negative' categories. The Softmax function is suitable for multi-class classification problems as it calculates the probability of each category.

The proposed model effectively captures contextual information through BERT embeddings and deeply learns features using BiLSTM and CNN layers. By emphasizing key words and contexts with the attention layer, customer response analysis accuracy was improved. The model demonstrated the ability to accurately classify text data generated from sales activities by customer response, showcasing its practical applicability in real-world B2B sales environments.

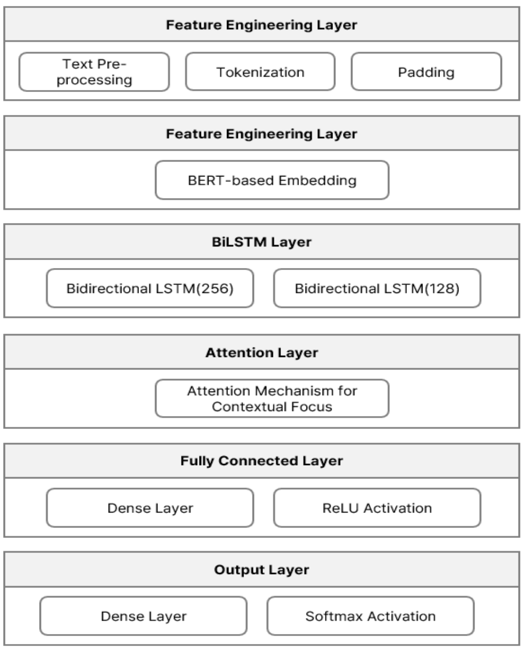

Fig. 1. Model Architecture

IV. Experiments and Results

This study used real-world data to conduct experiments, showing excellent performance and proving its practical applicability in future B2B sales environments.

A. Experimental Environment

In this study, unstructured text data such as meeting records, emails, and chat logs generated during sales activities were collected from a CRM system used in real-world operations. A total of 2,000 text samples were initially gathered. After removing irrelevant or system-generated data, 1,574 valid text samples were retained. The collected data were labeled into three classes: positive (124 samples), neutral (955 samples), and negative (495 samples). To address the data imbalance across these classes, data augmentation was applied, resulting in a balanced dataset of 2,865 samples.

Data preprocessing is an essential step to maximize the performance of NLP models. In this study, only Korean text was analyzed, and non-Korean characters and symbols were removed. Additionally, unnecessary stop words that do not contribute to sentiment analysis were eliminated. The KoNLPy (Korean Natural Language Processing in Python) library's Okt (Open Korean Text) morphological analyzer was utilized to segment the text data into morphemes for further analysis. Label encoding was performed to convert the sentiment classes (positive, neutral, negative) into numerical values, enabling efficient utilization of the data during model training.

Hyperparameters were set for model training. The learning rate was set at 0.001 to ensure stable convergence of the model, and the batch size was set at 16 to optimize memory usage. The number of epochs was set to 15 to provide sufficient training while preventing overfitting. Additionally, the dropout rate was set at 0.5 to mitigate overfitting.

B. Evaluation Methodology

Various metrics were used to evaluate the model's performance. **Accuracy** represents the proportion of correctly predicted samples out of all samples, serving as a basic indicator of the model's overall performance. **Precision** measures the proportion of positive predictions that are actually positive, evaluating the accuracy of the model's positive predictions. **Recall** indicates the proportion of actual positive samples correctly predicted by the model, focusing on the model's ability to avoid missing positive data. Finally, **F1 Score** is the harmonic mean of precision and recall, balancing the two metrics.

C. Results Analysis

Comprehensive analysis of the experimental results showed high performance for the model. Analysis of the confusion matrix revealed that the model correctly predicted 162 out of 186 samples in the positive class, 157 out of 194 samples in the neutral class, and 193 out of 193 samples in the negative class (100%). Analysis of the performance metrics showed an accuracy of 89.35%, a precision of 89.36%, a recall of 89.35%, and an F1 score of 89.30%. The model demonstrated high overall accuracy and balanced performance, particularly excelling in the negative class. This was attributed to the distinct patterns in negative response.

TABLE I. Experimental Results

Response	*Result*			
	Accuracy	*Precision*	*Recall*	*F1-score*
Positive	***0.89***	***0.83***	***0.87***	***0.85***
Neutral		***0.87***	***0.81***	***0.84***
Negative		***0.98***	***1.00***	***0.99***

TABLE II. Confusion Matrix of the Proposed model

Response		*Predicted*		
		Positive	*Neutral*	*Negative*
True	***Positive***	***0.83***	***0.87***	***0.85***
	Neutral	***0.87***	***0.81***	***0.84***
	Negative	***0.98***	***1.00***	***0.99***

D. Comparative Analysis

A comparison with the existing SCA-HDNN model showed that the proposed model achieved superior performance overall [8]. Analysis of the confusion matrix for the comparison model showed that it correctly predicted 193 out of 193 samples in the positive class, 136 out of 194 samples in the neutral class, and only 158 out of 186 samples in the negative class. Performance metrics for the comparison model showed an accuracy of 84.99%, a precision of 85.06%, a recall of 85.01%, and an F1 score of 84.59%.

In particular, while previous studies mainly used structured datasets, this study focused on unstructured data and showed excellent performance. In conclusion, the experimental results of this study showed better performance than the comparative model, and can make an important contribution, especially in that it has high potential for practical use in future B2B sales environments.

V. Conclusion

This study proposed a BiLSTM-based model that effectively classifies customer responses occurring in B2B sales activities and confirmed high accuracy and reliability. However, in future studies, it is necessary to expand the diversity of data, increase model interpretability, and conduct more sophisticated customer response analysis through multimodal data analysis. Ultimately, it is expected that this will maximize the efficiency of B2B sales activities and contribute to increasing corporate sales.

A. *Summary of the Study*

This study developed a customer response classification model using customer response text data collected from B2B sales activities and classified responses into positive, neutral, and negative categories. NLP techniques, such as stop word removal and morphological analysis, were applied during text preprocessing to prepare the data for learning. The proposed BiLSTM-based model was designed to learn bidirectional contextual information, and an attention layer was added to emphasize key information within the text sequence. The model's performance was optimized through learning rate scheduling and handling class imbalance [8].

B. *Limitations of the Study*

The limitations of this study can be summarized into three major points. First, the amount and variety of data are limited. The data used in the study were collected only in a specific B2B domain, and the generalizability of the model may be limited because it does not include data from various industries. Second, the customer responses were classified only as positive, no response, and negative, and thus the detailed differences in emotions were not reflected [9]. Finally, the interpretability of the model is limited, making it difficult to provide an explanation of how the model derives predictions in an actual application environment.

C. *Future Research Directions*

Future research should focus on collecting data from various B2B industries to enhance the model's generalization performance and analyzing various unstructured data sources to improve the accuracy of customer response analysis [10]. Additionally, interpretability techniques should be introduced to help sales representatives easily understand the model's predictions and incorporate them into decision-making [11]. Furthermore, integrating multi-modal analysis techniques that include text, voice, and video data could be a promising area of study.

Acknowledgment

This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2025-RS-2022-00156360)

References

[1] Kim Jin-Hwan, “The Role of Artificial Intelligence in B2B Sales and Marketing”, Department of Management of Technology Graduate School of Management of Technology Korea University, 2021.

[2] Ho Yeon Park, “Recommender Systems using Customer response Analysis with Deep Learning”, Department of Management Information Systems

[3] Olujimi, P. A., & Ade-Ibijola, A. “NLP techniques for automating responses to customer queries: A systematic review”, Discover Artificial Intelligence, 05, 2023.

[4] DataCamp Editorial Team, “Data Science in Sales: Customer Customer response Analysis”, DataCamp (Learn R, Python & Data Science Online), 2021.

[5] Aditya Pal, Aruna Malapati, “A BERT based Ensemble Approach for Customer response Classification of Customer Reviews and its Application to Nudge Marketing in e-Commerce”, arXiv, 2020.

[6] Rosaline Ningombam, Thoudam Doren Singh, “NLP Techniques for Automating Responses to Customer Queries: A Systematic Review”, Springer, 2021.

[7] Doru Rotovei, Viorel Negru, “Data driven sales prediction using communication customer response analysis in B2B CRM systems”, 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2019.

[8] Phillip Britt, “Content Is Key to Conversational AI Success”, destination CRM Magazine, September 2022.

[9] Shervin Minaee, Nal Kalchbrenner, et al., “Deep Learning Based Text Classification: A Comprehensive Review”, arXiv.org , 2020.

[10] Jawad Khan, Niaz Ahmad, Youngmoon Lee, Aftab Alam, “Leveraging Semantic and Customer response Knowledge for User-Generated Text Customer response Classification”, W-NUT, 2022.

[11] Heiko Fischer, Sven Seidenstricker, et al., “Artificial Intelligence in B2B Sales: Impact on the Sales Process”, Artificial Intelligence and Social Computing, Vol. 28, 2022.