Extracting Fine-Grained Sentiment Features about Library Services from Reader Feedback Text Using RoBERTa

Chenhui Song
Putian University Library, No. 2121 Zixiao East Road, Xitianwei Town, Licheng District, Putian 351100, Fujian, China
Corresponding author: Chenhui Song (e-mail: 15980353836@163.com)

ABSTRACT Robustly Optimized Bidirectional Encoder Representations from Transformers Approach (RoBERTa) model, fine-tuned for domain adaptation with library domain data, and introducing the Aspect-Based Sentiment Analysis (ABSA) framework to extract and classify fine-grained sentiment related to collection services in reader feedback. Based on annotation and experiments with 23,684 open reader feedback texts from 12 university libraries in China over the past three years, a fine-grained annotation system is constructed, encompassing six collection service dimensions, three types of sentiment polarity, including positive, neutral, and negative, and three levels of intensity. The results show that the domain-fine-tuned RoBERTa model achieved an F1 score of 86.7% in aspect recognition and an accuracy of 89.3% in sentiment polarity classification. The aspect recognition F1 score is significantly better than baseline models such as Bidirectional Encoder Representations from Transformers (BERT) by 5.2% and Text Convolutional Neural Network (TextCNN) by 11.6%. Among the evaluated dimensions, access to electronic resources and speed of obtaining books are the two dimensions with the highest concentration of negative sentiment, accounting for 31.4% and 27.8% of the total negative feedback, respectively. This research method improves the reliability and interpretability of sentiment analysis in professional service scenarios and provides a data-driven reference for optimizing digital resource services for engineering disciplines involving electromagnetic theory, antenna technology, and propagation research.

INDEX TERMS roberta, library collection services, aspect-based sentiment analysis, reader feedback, electromagnetic engineering

I. INTRODUCTION

WITH the rapid development of information technology, the global library industry is undergoing a profound digital transformation, which is also driving the need for better access to specialized digital archives and research resources relevant to advanced engineering disciplines. Library resources are gradually shifting from paper to digital carriers, and service models are also being comprehensively upgraded. Readers' needs are no longer limited to tradi-tional document borrowing, but have expanded to digital resource acquisition, intelligent consultation, re-mote access, personalized recommendations, and other aspects; service scenarios have also extended from a single offline space to a hybrid form that integrates online and offline [1,2]. In this process, the volume of reader feedback has increased dramatically [3]. Sources include library official website message boards, mo-bile application reviews, social media discussions, reader symposium records, and other channels, forming a large-scale textual data resource [4]. These data truly reflect readers' views on collection quality, service efficiency, space experience, and technical support [5], and have become a key basis for libraries to optimize services and improve reader satisfaction [6]. Faced with large volumes of unstructured text, the efficient extraction of valuable information and the transformation of libraries from data sedimentation to intelligent services [7,8] have become a core issue in the process of digital transformation.

Although libraries have accumulated a large amount of reader feedback data, they are currently facing a prominent contradiction between data abundance and insufficient insights. On the one hand, unstructured text data accounts for the vast majority of feedback information, which contains a large number of fragmented opinions, emotional expressions and professional terminology [9]. Traditional manual analysis methods are limited by efficiency and subjectivity and are difficult to achieve large-scale processing [10]. On the other hand, existing general sentiment analysis models mainly judge the overall sentiment polarity (such as positive, negative, and neutral) [11] and lack the ability to parse specific service dimensions in a fine-grained manner [12]. For example, when a reader reports that "the library's electronic resources are updated slowly but the staff's service attitude

is very good", a general model can only identify the overall mixed sentiment [13] but cannot distinguish between the two different sentiment tendencies: electronic resource updates (negative) and staff service attitude (positive). This limitation makes it difficult for library managers to accurately identify service shortcomings [14,15] and cannot achieve quantitative evaluation of various dimensions of collection services, making it difficult for valuable feedback data to fully realize its decision-making support value.

In recent years, the development of sentiment analysis technology has provided new possibilities for resolving the above contradictions, and related research has shown multi-dimensional progress in different fields. In the field of higher education, there have been preliminary explorations into applying academic sentiment analysis to teaching feedback processing [16]. For example, through simple keyword matching or basic machine learning models to identify sentiment tendencies in feedback texts [17,18], qualitative references can be provided for service optimization. However, these attempts generally remain at the level of overall sentiment judgment and have not yet achieved fine-grained division of service aspects. At the same time, ABSA [19] has shown significant advantages in practice in other professional fields, such as successfully identifying sentiment polarity in specific aspects such as product quality and logistics speed in the e-commerce field [20], and accurately extracting user reviews of dimensions such as taste, environment, and service in the catering industry [21]. This approach deepens sentiment analysis from the overall level to the attribute level and establish a direct mapping between specific aspects and sentiment polarity, thereby enabling fine-grained sentiment analysis [22,23]. On the technical level, pre-trained language models represented by RoBERTa [24] have surpassed traditional models in text comprehension capabilities by optimizing pre-training strategies [25]. To effectively identify terminology and complex semantics in library professional texts, the RoBERTa model provides a more powerful technical foundation than BERT [26]. In the pre-training phase, it achieves better semantic understanding and generalization capabilities by removing next-sentence predictions, expanding the data scale, and introducing dynamic masking. After domain fine-tuning, it can accurately serve the professional scenario analysis of libraries.

This study aims to develop a fine-grained sentiment analysis framework for library collection services. To accurately quantify specific service aspects and sentiment in reader feedback, the study employed a technique that combines domain-adaptive RoBERTa with ABSA. First, RoBERTa is fine-tuned using domain-specific corpora to enhance its understanding of specialized terminology. Second, a multi-task learning model is constructed to simultaneously perform aspect recognition and sentiment classification, improving the accuracy of the overall analysis. Experimental results demonstrate that this method achieves an F1 score of 86.7% for the collection service aspect recognition task and an accuracy of 89.3% for the sentiment polarity classification task. This method addresses the data-rich but insight-poor limitation of traditional approaches by providing an interpretable and quantifiable analytical tool for the multi-dimensional evaluation and precise optimization of library collection services.

II. CONSTRUCTION OF FINE-GRAINED SENTIMENT ANNOTATION SYSTEM AND DATA PREPARATION FOR COLLECTION SERVICES

DATA SOURCE AND COLLECTION

The reader feedback text collection work in this study covers 12 university libraries in China, including representative universities such as comprehensive, science and engineering, and teacher training. The time span is nearly three years (2022-2024), and the data is timely and representative.

Data were collected from two types of feedback channels: online and offline.

Online channels included the library's official website message board, WeChat public account backend messages, email feedback, the mobile library APP review section, and campus Bulletin Board System (BBS) discussion posts.

Offline channels included paper suggestion box submissions, on-site consultation records, and reader symposium minutes.

A total of 23,684 valid reader feedback texts were collected, forming a text corpus covering the entire scene of library collection services. All texts are desensitized to remove personal information such as readers' names, student numbers, contact information, etc., retaining only semantic content, in accordance with scientific research ethics and data security standards. The data source distribution statistics are shown in Table 1.

TABLE 1. Data source distribution statistics

Library Type	Number of Institutions	Time Span	Total Valid Feedback Entries
Comprehensive	5	2022–2024	10,245
Science & Engineering	4	2022–2024	8,732
Normal (Teacher Training)	3	2022–2024	4,707
Total	12	2022–2024	23,684

PREPROCESSING PROCESS

To ensure the quality of the annotated data, the original text needs to be preprocessed in multiple steps. The specific process is as follows:

Deduplication is performed using the text fingerprint algorithm SimHash [27] to calculate the similarity of the collected text, setting a threshold of 0.95 to remove redundant content that is repeated or highly similar (such as duplicate complaints copied and pasted), and retain text units with independent information value. In the irrelevant information filtering stage, irrelevant content is removed by combining keyword matching with a rule engine, including personal identity information (such as mobile phone numbers, student numbers), advertising and promotion information, complaints about other campus facilities (such as cafeteria and dormitory problems), and

meaningless emotional expressions (such as simple insults or emoticons). Text standardization focuses on resolving inconsistent terminology and colloquial expressions: on the one hand, a professional term mapping table is established to standardize synonyms (such as "database", "electronic resource library", and "document library") into "electronic resource database", and "borrowing and returning books" and "borrowing" into "borrowing service"; on the other hand, colloquial expressions are converted into written form, such as standardizing "unable to borrow books" into "insufficient availability of book borrowing", and adjusting "the network is too slow" to "poor network access speed". At the same time, special symbols (such as #, @, emoticons) and redundant punctuation in the text are removed to ensure the standardization and consistency of text expression.

EMOTIONAL DIMENSION DESIGN

Based on the core scenarios of collection services and the high-frequency themes of reader feedback, this study constructed a service classification system consisting of six first-level dimensions. The definitions of each dimension are shown in Table 2.

Table 2 details the description of the service classification system and the polarity and intensity of sentiment annotations for the six first-level dimensions. The six dimensions are:

(1) Book acquisition speed: refers to the efficiency of the entire process of the library's newly published documents (including books, journals, dissertations, etc.) from procurement application, cataloging and processing to shelving and circulation. The core evaluation indicators include the new book shelving cycle, the coverage of popular textbooks, and the timeliness of subject documents;

(2) Electronic resource access: focuses on the quality of digital resource services, covering technical service elements such as database access stability (such as login success rate), off-campus access permission configuration, full-text download speed, resource update frequency, and ease of use of search functions;

(3) Borrowing services: This involves the entire circulation service process, including the borrowing period setting, the convenience of renewal operations, the efficiency of reservation notification, the overdue handling policy, and the borrowing rules for special collections (such as ancient books and special collections);

(4) Reading environment: This focuses on the physical space experience, including the division of library areas (such as quiet areas and discussion areas), seat comfort, lighting and ventilation conditions, temperature control, noise level, and cleaning and maintenance conditions;

(5) Consulting services: This measures professional service capabilities, including the speed of reference consultation response, the depth of subject librarian services, the quality of information literacy training, and the degree of satisfaction of personalized needs;

(6) Facility maintenance: This involves the operating status of hardware equipment, including the failure rate of self-service equipment such as self-service borrowing and returning machines, printers, and copiers, the quality of network coverage, and the response time for fault reporting.

Sentiment polarity is categorized as positive, neutral, and negative. Positive sentiment indicates clear satisfaction or positive evaluation of the service, typically expressed as, "Access to electronic resources is extremely fast, and off-campus access is also convenient. I'm very satisfied". The neutral category does not represent a specific emotional state, but rather refers to service-related statements that do not contain explicit evaluations or emotional cues. It typically manifests as an objective description of the service usage process, conditions, or phenomena. In the labeling system, this category serves as the baseline state for emotional polarity and is used to distinguish between feedback texts that contain evaluative content and that do not, such as "The library added 500 new humanities books this week".

In Table 2, although "electronic resource access" and "facility maintenance" have a causal relationship in real-world scenarios (e.g., network coverage failure may lead to access failure), they correspond to different analytical levels in this study. "Electronic resource access" focuses on the digital resource services from the reader's perspective, including successful database login, completion of off-campus access configuration, and full-text retrieval, whereas "facility maintenance" focuses on the infrastructure supporting these services, such as network coverage quality, the stability of self-service equipment, and the response time for fault repair.

In the sentiment labeling process, this study follows the labeling principle of "primary cause priority and clear service orientation": when feedback emphasizes the availability or access result of the resource service itself (even if caused by network problems), it is labeled as "electronic resource access"; when the feedback mainly describes infrastructure malfunction or delayed maintenance without referring to specific resource access behavior, it is labeled as "facility maintenance". This differentiation strategy enables the model to capture two complementary sentiment signals—service experience differences and infrastructure reliability—within a multi-task learning framework.

ANNOTATION GUIDELINES

This study employs a systematic three-stage annotation process: independent annotation, cross-verification, and dispute resolution. Specifically, the pre-processed text is evenly distributed among three professionally trained annotators, who independently complete three core tasks: determining the service dimension to which the text belongs (multiple choices allowed), determining sentiment polarity (single choice), and assessing sentiment intensity (single choice), while also recording the annotation basis (such as keywords and contextual clues). After annotation, the system automatically compares the three annotators' results. If

TABLE 2. Definition of fine-grained sentiment annotation dimensions for collection services

Aspect Dimension	Definition	Sentiment Polarity	Intensity Criteria
Book Acquisition Speed	Efficiency of the entire workflow from procurement request to shelf availability for newly published materials	Positive: Praise for timely updates. Neutral: Factual statement. Negative: Complaints about delays.	Mild: Occasional delay, minor impact. Moderate: Frequent delays affecting specific user groups. Strong: Systemic, long-term unavailability impacting all users.
Electronic Resource Access	Quality of digital resource services, including database login stability, off-campus access configuration, full-text download speed, update frequency, and usability of search functions	Positive: Smooth access experience. Neutral: Factual or descriptive statements related to the service but not containing explicit evaluation or emotional bias. Negative: Access failures or poor performance	Mild: Temporary glitches. Moderate: Recurrent issues in specific resources. Strong: Prolonged system outage affecting all users.
Circulation Services	End-to-end borrowing process, including loan periods, ease of renewal, notification for reserved items, overdue policies, and rules for special collections	Positive: Flexible and user-friendly policies. Neutral: Policy description without sentiment. Negative: Frustration with restrictions or system errors.	Mild: One-time renewal error. Moderate: Short loan periods hindering research. Strong: Inability to borrow critical materials due to rigid rules.
Reading Environment	Physical space experience, covering zoning (quiet/study/collaboration areas), seating comfort, lighting, ventilation, temperature, noise levels, and cleanliness	Positive: Comfortable and well-maintained spaces. Neutral: Objective description of layout. Negative: Complaints about discomfort or poor maintenance.	Mild: Minor inconvenience. Moderate: Widespread issues. Strong: Environment unsuitable for study.
Reference & Consultation Services	Professional support quality, including response speed of reference desk, depth of subject librarian assistance, effectiveness of information literacy training, and fulfillment of personalized requests	Positive: Prompt and helpful service. Neutral: Mention of service availability. Negative: Unresponsive or superficial support.	Mild: Slight delay in reply. Moderate: Inadequate help for complex queries. Strong: Complete lack of specialist support for research needs.
Facility Maintenance	Operational status of physical infrastructure and self-service equipment, including reliability of self-checkout machines, printers, copiers, network coverage quality, and responsiveness of repair services	Positive: Equipment works reliably. Neutral: Factual mention. Negative: Frequent breakdowns or slow repairs.	Mild: Occasional device glitch. Moderate: Recurrent issues with key equipment. Strong: Widespread outages or unresponsive maintenance affecting core services.

all three criteria are consistent, the result is adopted. If any disagreement arises (such as inconsistent dimension determination, polarity, or intensity), a review mechanism is initiated, in which the three annotators review the text context and discuss and negotiate according to the annotation guidelines. If consensus could not be reached, the result was submitted to a panel of annotation experts consisting of two library science experts and one natural language processing researcher for final arbitration, ensuring the accuracy and consistency of the annotation results.

To ensure the quality of annotation, the study implements multi-stage, multi-dimensional quality control measures. Before annotation, a two-week systematic training program is organized. This training focuses on standardizing annotation knowledge through theoretical explanations (covering service dimension definitions, sentiment classification criteria, and more) and case studies (exercises with 100 sample texts). During the annotation process, a random sampling of 5% of completed texts is conducted weekly, with an expert panel assessing accuracy and providing timely feedback on improvements. A dynamic update mechanism is also established to promptly summarize and supplement new issues that arose during the annotation process into the annotation manual, ensuring continuous optimization of annotation standards.

III. MULTI-TASK FINE-GRAINED SENTIMENT ANALYSIS MODEL BASED ON ROBERTA

OVERALL ARCHITECTURE DESIGN

The fine-grained sentiment analysis model constructed in this study uses the pre-trained RoBERTa model as its core encoder. It enables in-depth analysis of reader feedback through a basic encoding-multi-task offload architecture. The model consists of three functional modules—a text encoding layer, a feature sharing layer, and a multi-task output layer—forming a complete processing chain of input-encoding-offload-prediction. The overall model architecture is shown in Figure 1.

The RoBERTa encoder, as the basic component of the model, is responsible for converting raw text into a vector representation rich in semantic information. Taking the input of "Off-Campus access to electronic databases often crashes, I can't log in at all when writing papers at home" as an example, the encoder uses a 12-layer Transformer structure. The core of each layer is a multi-head self-attention mechanism, which is the key to capturing fine-grained semantics.

The self-attention mechanism calculates the correlation weight between any two tokens in the input sequence and dynamically aggregates contextual information [28]. For a token in the l-th layer, its representation is updated as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \qquad (1)$$

In formula (1), Q (Query), K (Key), and V (Value) are all derived from linear transformations of the token representations in the previous layer, and d_k represents the scaling factor.

The multi-task output layer consists of two parallel task heads: an aspect classification head and a joint sentiment polarity and intensity prediction head. The former uses token-level features to identify specific aspects of the text related to collection services; the latter uses a comprehensive judgment based on tokenlevel and sentence-level features to accurately identify the sentiment polarity (positive/negative/neutral) and intensity (mild/moderate/strong) associated with specific aspects. These two task heads share information by sharing RoBERTa's underlying features and are jointly optimized during training, improving the model's ability to capture fine-grained sentiment features.

DOMAIN ADAPTIVE FINE-TUNING STRATEGY

Pre-training is continued on library corpora, using text corpora from library websites, Online Public Access Catalogues (OPAC) systems, relevant papers, and other fields, using the Masked Language Modeling (MLM) task. The specific strategy is shown in Figure 2.

In order to adapt the general pre-training model to the characteristics of library domain text, domain adaptive fine-

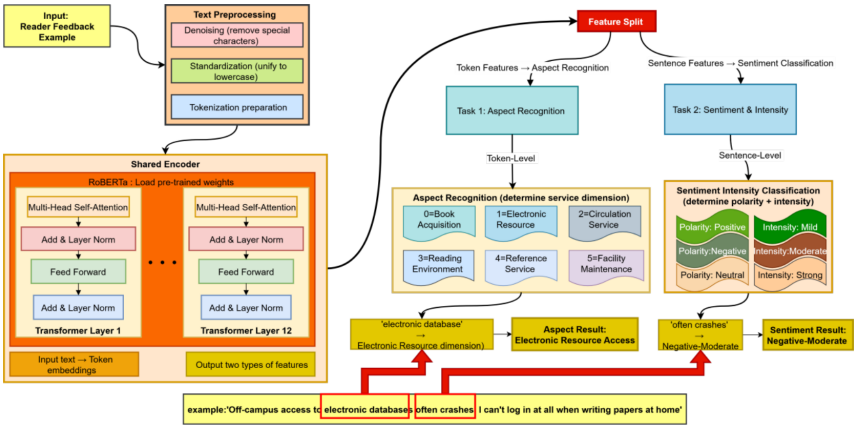

FIGURE 1. Overall Model Architecture

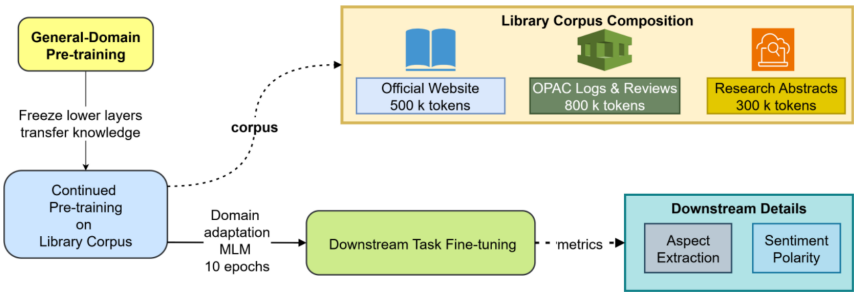

FIGURE 2. Domain Adaptive Fine-tuning Process

tuning is required. As shown in Figure 2, by continuing pre-training on library professional corpus, the model's ability to understand domain terminology, service scenarios and user expression habits is enhanced.

LIBRARY DOMAIN CORPUS CONSTRUCTION

The construction of the corpus follows the principles of multi-source integration and professional focus, covering three core data sources:

Library website texts include official documents such as collection service introductions, borrowing rules, and reader guides, totaling about 500,000 words, to ensure that the model masters standard service terminology; OPAC system [29] interaction data includes user search records, resource evaluation messages, and consultation dialogues collected from the library's online public access catalog, totaling about 800,000 words, covering natural language expressions in users' actual usage scenarios;

Library science research literature selects abstracts and keywords of collection service-related papers included in China National Knowledge Infrastructure (CNKI) [30], totaling about 300,000 words, to supplement field professional vocabulary and academic expressions.

After preprocessing, all texts form a corpus with a total size of approximately 1.6 million words. Data quality is ensured through deduplication, removal of special symbols, and standardization (e.g., unifying synonyms such as "borrowing" and "borrowing and returning").

CONTINUE PRE-TRAINING TASKS AND PARAMETER SETTINGS

Domain adaptive fine-tuning uses MLM [31] as the pre-training objective. This process simulates the clozestyle language comprehension mechanism by randomly masking 15% of the tokens in the input text (e.g., replacing "self-service lending machine" in "self-service lending machine operation is convenient" with [MASK]) and training the model to predict the masked content based on the context. This task can enable the model to learn the semantic associations and expression rules of the library domain, rather than simply memorizing the text.

The training rounds are set to 10 to balance model convergence and overfitting risks. The perplexity on the validation set is used to monitor the training progress, and the training is terminated early if the perplexity did not decrease after three consecutive rounds. The initial learning rate is $5\times10\text{-}5$, which is lower than the common pre-training stage (10-4) to avoid drastic parameter updates that would destroy the original language knowledge. The batch size is set to 32, and gradient accumulation technology is used to simulate the effect of larger batch training within the memory limitations of a single Graphics Processing Unit (GPU). For core semantic tokens such as nouns and verbs, a strategy of replacing them with [MASK] with 80% probability, randomly replacing them with other tokens with 10% probability, and leaving them unchanged with 10% probability is adopted to enhance model robustness.

MULTI-TASK LEARNING FRAMEWORK

The multi-task learning framework achieves end-to-end analysis of aspect recognition and sentiment assessment by jointly optimizing two closely related subtasks. The two subtasks mutually enhance each other during training, improving overall analysis accuracy. At the token level, the Aspect Branch uses a Bidirectional Long Short-Term Memory (BiLSTM) softmax algorithm to assign Begin, Inside, Outside (BIO) aspect labels to each word. The Sentiment Branch represents the entire sentence using a [CLS] vector. The model outputs joint probabilities for nine polarity–intensity categories through two fully connected layers, enabling simultaneous aspect recognition and sentiment assessment. As shown in the multi-task learning framework in Figure 3, the two branches each calculate their own losses (cross entropy for aspect and weighted cross entropy for sentiment), which are then weighted and combined to form the total loss Ltotal at a ratio of 1:1.2. Backpropagation simultaneously updates the shared encoder and each task head.

ASPECT CATEGORY RECOGNITION TASK

The BIO annotation system is used. In Figure 3, B-Aspect represents the starting word of an aspect entity, I- Aspect represents the middle word of an aspect entity, and O represents a non-aspect word. Each token in the text is classified into six categories. The model uses the token-level vector output by the feature sharing layer to connect to a BiLSTM network to capture local contextual dependencies. Classification probabilities are then output through a linear layer and a softmax function.

The loss function uses cross entropy loss. To address the problem of class imbalance, it applies class weights (the weight value is inversely proportional to the sample frequency) and formulate it as follows:

$$L_{\text{aspect}} = -\sum_{i=1}^{N} w_{c_i} \log(p_{c_i}) \tag{2}$$

In formula (2), w_{c_i} is the category weight, and p_{c_i} is the probability that the model predicts that the i-th token belongs to category c_i.

JOINT PREDICTION OF SENTIMENT POLARITY AND INTENSITY

After identifying aspect entities, the sentiment polarity (positive/negative/neutral) and intensity (mild/moderate/strong) of the aspect must be determined. This is a sentence-level multi-label classification task. A two-dimensional polarity-intensity labeling system is used to classify sentiment into nine fine-grained categories (3 polarities x 3 intensities). The model uses the [CLS] special token vector (representing the semantic meaning of the entire sentence) output by the feature sharing layer, passing it through a two-layer fully connected network with a GELU activation function to output predicted probabilities for each of the nine categories.

$$L_{\text{sentiment}} = -\sum_{j=1}^{9} w_j \cdot y_j \cdot \log(p_j) \tag{3}$$

In formula (3), j is the index of the sentiment category, which has nine categories; y_j represents the one-hot

encoding of the true label; p_j is the model's predicted probability of category j, satisfying $\sum_{j=1}^{9} p_j = 1$; and w_j is the weighting coefficient for category j, which is used to alleviate category imbalance and enhance intensity differentiation. Specifically, the intensity level weights are set to 1.0 for Mild, 1.2 for Moderate, and 1.5 for Strong. A weighted multi-class cross-entropy loss is used, with increasing weights assigned to the intensity dimension to enhance the model's ability to distinguish differences in sentiment intensity.

FEATURE SHARING AND JOINT OPTIMIZATION MECHANISM

The two subtasks employ hard parameter sharing to reuse features. The RoBERTa encoder and feature sharing layer parameters are fully shared across tasks, with only the task-specific heads being independently trained. This architecture substantially reduces the parameter count while enabling the model to learn more generalized domain feature representations.

The joint loss function is calculated as follows:

$$L_{\text{total}} = L_{\text{aspect}} + 1.2 \times L_{\text{sentiment}} \tag{4}$$

In formula (4), the losses of the two tasks are weighted and summed with a ratio of 1:1.2 (the sentiment prediction task has a higher weight due to its finer-grained labels) as the overall training objective. During training, the shared parameters and task head parameters are updated simultaneously through backpropagation, allowing the model to be collaboratively optimized for both aspect recognition and sentiment prediction.

IV. EXPERIMENT AND RESULT ANALYSIS

A. EXPERIMENTAL SETUP

The selection of baseline models is based on classic and cutting-edge methods in the field of natural language processing, forming a multi-level comparison system: BERT, as a representative pre-trained language model, has a bidirectional attention mechanism that can capture contextual semantic associations and is widely used as a benchmark for text classification tasks [32,33]; TextCNN extracts local n-gram features through convolutional layers and is suitable for short text sentiment analysis [34]; BiLSTM-Attention combines the sequence modeling capabilities of recurrent neural networks with the key information focusing capabilities of the attention mechanism and is good at processing the temporal dependencies of emotional expressions [35]. The three baseline models represent the three major technical routes of pre-trained language models, convolutional neural

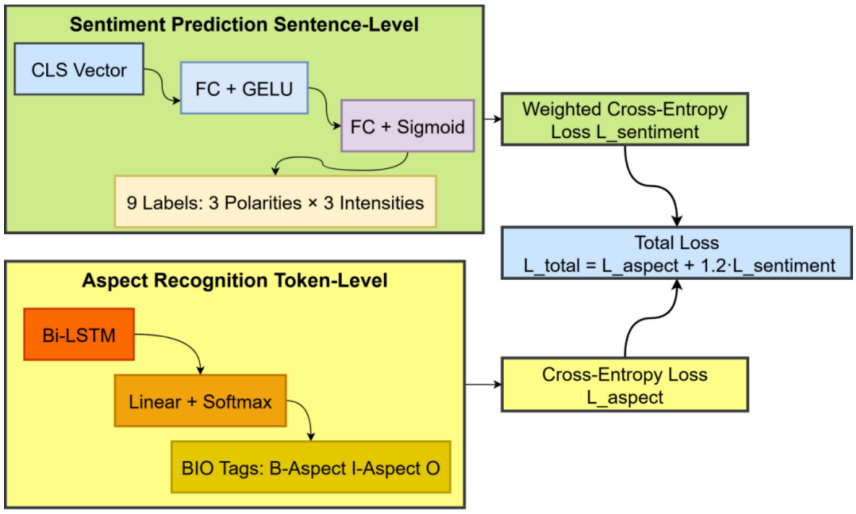

FIGURE 3. Multi-task learning framework

networks, and recurrent neural networks. Through comparison, the technical advantages of the RoBERTa model in fine-grained sentiment analysis tasks can be clearly seen.

The evaluation metrics use three core indicators from the field of sentiment analysis. Accuracy is defined as the proportion of samples correctly classified by the model to the total number of samples, reflecting the overall classification performance. F1-Score is the harmonic mean of Precision and Recall, calculated as 2 × (Precision × Recall) / (Precision + Recall), which is used to balance the model's ability to distinguish between positive and negative samples. Macro-F1 calculates the F1 score for each sentiment category and then takes the arithmetic average to eliminate the impact of category imbalance on the evaluation results. All models in the experiment are trained under the same hardware environment (NVIDIA Tesla V100 GPU) and software framework (PyTorch 1.10). The hyperparameters of the RoBERTa model are set as follows: learning rate 2×10-5, batch size 32, training epochs 10, weight decay 10-4, and the AdamW optimizer with a linear learning rate warmup strategy. The hyperparameters of the baseline model are determined after grid search optimization. The learning rate of BERT is 2×10-5, the convolution kernel size of TextCNN is 2, and the hidden layer dimension of BiLSTM-Attention is 256.

MAIN RESULTS

The model performance comparison results are shown in Table 3.

TABLE 3. Model Performance Comparison

Model	Aspect Identification F1 (%)	Sentiment Classification Accuracy (%)	Macro-F1 (%)
RoBERTa	86.7	89.3	85.9
BERT-base	81.5	84.5	80.7
TextCNN	75.1	76.1	73.4
BiLSTM-Attention	78.4	79.6	76.8

Table 3 presents a performance comparison indicating that the RoBERTa model achieves optimal performance in the fine-grained sentiment analysis task for collection services, achieving an F1 score of 86.7% for aspect recognition, 89.3% for sentiment classification, and 85.9% for macro-average F1. Compared to baseline models, RoBERTa achieves significant improvements in key metrics: Its aspect recognition F1 score increases by 5.2 percentage points compared to the BERT model (81.5%), by 11.6 percentage points compared to TextCNN (75.1%), and by 8.3 percentage points compared to BiLSTM-Attention (78.4%). Sentiment classification accuracy also increases by 4.8 percentage points compared to BERT (84.5%), by 13.2 percentage points compared to TextCNN (76.1%), and by 9.7 percentage points compared to BiLSTM-Attention (79.6%). RoBERTa optimizes its ability to capture contextual semantics during pre-training through a dynamic masking mechanism, a larger training corpus, and longer training time. In fine-grained tasks, its deep bidirectional Transformer architecture is able to better identify domain terms related to specific service dimensions, such as electronic resource access and book acquisition speed, and distinguish between neutral and negative sentiment intensity. This is the primary reason for its performance advantage over BERT's static masking, TextCNN's local feature limitations, and BiLSTM-Attention's inadequate long-distance dependency modeling.

ABLATION EXPERIMENT

Ablation experiments are conducted using the same dataset and training configuration, with only minor modifications to the model architecture or initialization method to isolate the impact of key modules. The specific design included three sets of control experiments:

The first set removes the pre-trained weights and replaces the RoBERTa encoder with a Transformer encoder of the same architecture (number of layers, hidden dimensions, etc.) but randomly initialized, resulting in a model without pre-training.

The second set disables the attention mechanism by re-

placing all multi-head self-attention layers in the RoB- ERTa encoder with simple feed-forward connections, retaining only the positional encoding and feed-forward network, eliminating the model's ability to dynamically focus on key information. This resulted in a model without the attention mechanism.

The third set eliminates the multi-task joint training framework and split aspect recognition and sentiment classification into two independent single-task models, each with its own encoder and task head, to examine the effectiveness of inter-task collaborative optimization. The ablation analysis is shown in Figure 4.

The experimental results in Figure 4 demonstrate that each component significantly impacts model performance.

In the no-pre-training setting, the F1 score for the aspect recognition task dropped significantly to 68.2%, 18.5 percentage points lower than the full RoBERTa model (86.7%). This demonstrates that the deep linguistic knowledge acquired by RoBERTa through large-scale corpus pretraining, including lexical semantics, syntactic structure, and contextual reasoning, is crucial for understanding library professional contexts. Models without pre-training struggle to generalize, and are particularly vulnerable to fine-grained tasks with limited labeled data.

The single-task model performed significantly worse than the multi-task framework, with the aspect recognition F1 decreasing to 83.1% and the sentiment classification accuracy to 86.4%. This result demonstrates the semantic complementarity between the two tasks. Aspect recognition provides a contextual anchor for sentiment judgment, while sentiment signals in turn guide aspect boundary detection.

ANALYSIS OF SENTIMENT DISTRIBUTION OF COLLECTION SERVICES

Based on the prediction results of the RoBERTa model, a quantitative analysis of the sentiment distribution of each dimension of collection services is conducted. It is found that the sentiment distribution of each dimension of collection services showed obvious differentiation characteristics, and negative emotions showed significant dimensional concentration characteristics. The sentiment distribution heat map is shown in Figure 5.

The heat map in Figure 5 shows that negative sentiment is concentrated in the Book Acquisition Speed (31.4%) and Electronic Resource Access (27.8%) dimensions, with the proportion of negative sentiment in these two dimensions significantly higher than in other service dimensions, reflecting that the timeliness of new book acquisition and the accessibility of electronic resources are currently the most significant user pain points. Positive sentiment is most prominent in the Circulation Services (50.9%) and Reading Environment (46.9%) dimensions, representing the specific service areas with the highest user satisfaction, indicating that optimizing the borrowing process and developing a good reading environment have become core service advantages of libraries.

Neutral sentiment is relatively evenly distributed across all dimensions but is highest in Facility Maintenance (58.5%) and Reference & Consultation Services (50.4%), reflecting a positive user attitude toward these services, with most users indicating that there is still room for improvement.

Further analysis focuses on the polarity of negative sentiment (mild/moderate/strong). The polarity values of negative sentiment for the six service dimensions are shown in Figure 6.

Figure 6 focuses on the intensity of negative sentiment across six core library service dimensions (measured as an average polarity value on a scale of -1 to 1, with values closer to -1 indicating stronger negative feedback). "Access to electronic resources" (-0.73) and "Speed of obtaining books" (-0.70), shown in dark red, represent the two dimensions with the highest proportions of negative sentiment. Both dimensions have significantly lower values than the other dimensions, indicating that users' most prominent complaint is the lack of timely access to resources.

The red dotted line represents the overall average negative sentiment of all service dimensions (-0.49). Except for "access to electronic resources" and "speed of obtaining books", which are far below this average (the degree of negativity far exceeds the overall level), the negative intensity of the other four dimensions, "circulation service" (-0.35), "reading environment" (-0.30), "reference consultation" (-0.45), and "facility maintenance" (-0.40) are all closer to or higher than the average, indicating that users' negative feedback in these dimensions is relatively mild.

ANALYSIS OF TYPICAL CASES

Three typical cases are used to demonstrate the model's capture of fine-grained emotional features. The attention weights (importance scores, not normalized) of the three cases are visualized in Figure 7.

The visualization of the attention weight distribution in Figure 7 demonstrates its ability to capture finegrained sentiment. In the case of negative electronic resource access: "Campus network login succeeded, but the database still can't open, delayed my thesis data download" (Figure 7(a)), the model assigns 87.6% of the attention weight to the keyword "can't open" and 78.5% to "delayed". "Can't open" directly points to the service malfunction, while "delayed" reinforces the impact of the negative outcome. Ultimately, the model correctly identified the case as "Electronic Resource Access–Negative". In this case, although the model's attention weight for "but" (3.2%) is relatively low, it confirms the persistence of the issue through contextual association, demonstrating its deep semantic understanding capabilities.

In the negative case study "Library has no new books for three months, popular novel section is still empty" (Figure 7(b)), attention is focused on the time adverbial "three months" (weight 42.1%) and the state description "empty" (weight 38.5%). "Three months" quantifies the ex-

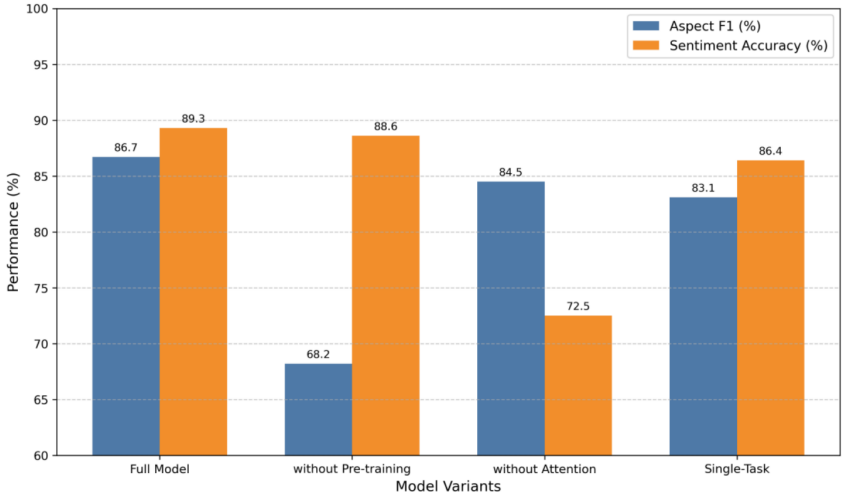

FIGURE 4. Ablation results analysis

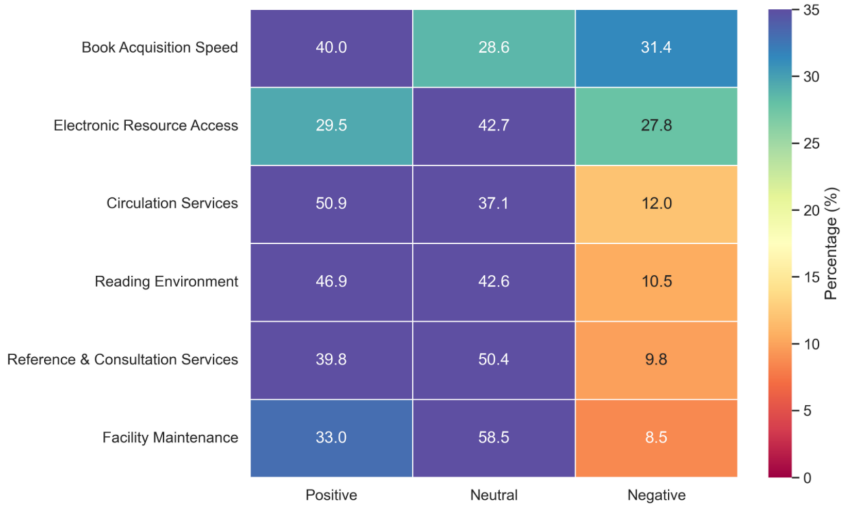

FIGURE 5. Distribution heatmap of positive/neutral/negative emotions in six dimensions.

tent of update delays, while "empty" embodies the resource shortage, successfully determining the sentiment polarity as strongly negative. In the positive case study "Circulation Services", "Self-service kiosk is simple, borrowing takes only 30 seconds, much faster than a manual counter" (Figure 7(c)), attention is focused on "simple" (weight 29.3%), "30 seconds" (weight 35.7%), and "faster" (weight 27.5%). The quantitative description of "30 seconds" and the comparative form of "faster" identify the sentiment of "borrowing service-positive". These cases demonstrate that RoBERTa's attention mechanism can adaptively focus on core vocabulary expressing sentiment and integrate contextual semantics to achieve accurate classification of fine-grained sentiment.

SENSITIVITY EXPERIMENT OF JOINT LOSS WEIGHTS

This paper conducts a systematic sensitivity analysis of the weights of the sentiment task relative to the aspect task in the joint loss. The experimental steps are as follows: while keeping other hyperparameters constant, a grid search is performed on the set {1.0, 1.1, 1.2, 1.5, 2.0} for the weights of the sentiment task (weight = 1 relative to the aspect task). For each setting, 5-fold cross-validation on the training set is used and repeated 3 times with random seeds to reduce randomness. The Aspect Identification F1, Sentiment Classification Accuracy, and Macro-F1 are recorded for each fold. Performance is reported as mean ± standard deviation, and Macro-F1 is used as the main indicator for weighing the overall performance of aspect recognition and sentiment classification. The results of the sensitivity experiment of the joint loss weights are shown in Table 4.

TABLE 4. Results of the Sensitivity Experiment of Joint Loss Weights

Emotional task weight	Aspect Identification F1 (%)	Sentiment Classification Accuracy (%)	Macro-F1 (%)
1.0	86.0 ± 0.9	88.3 ± 0.7	85.2 ± 0.8
1.1	86.4 ± 0.8	88.9 ± 0.6	85.6 ± 0.7
1.2	86.7 ± 0.6	89.3 ± 0.5	85.9 ± 0.5
1.5	85.9 ± 0.7	89.5 ± 0.6	85.3 ± 0.6
2.0	85.1 ± 1.0	89.7 ± 0.8	84.7 ± 0.9

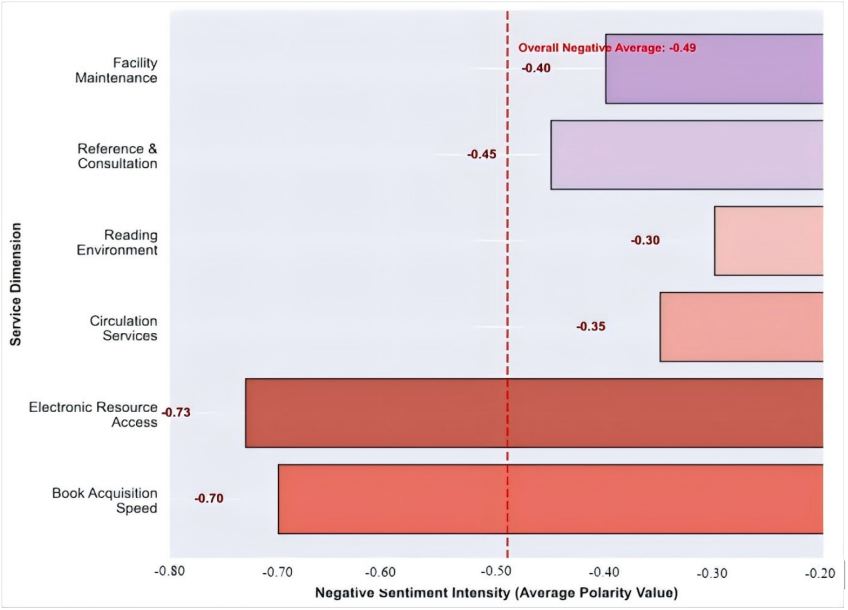

FIGURE 6. Negative emotion polarity on six dimensions

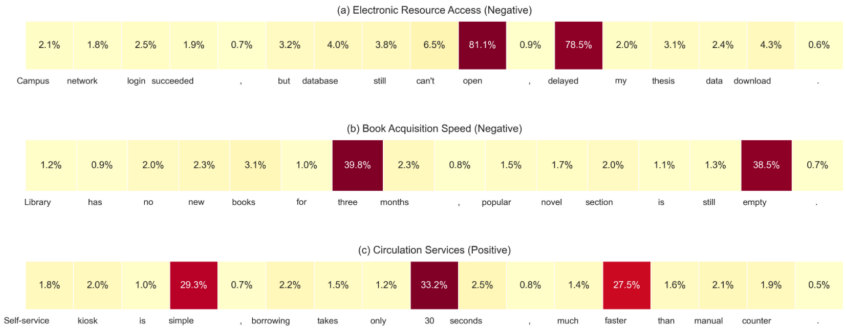

FIGURE 7. Attention weight visualization example. The red box represents the weight ratio, and the redder the color, the greater the weight value

The results in Table 4 show that the weight of the sentiment task in the joint loss has a significant impact on the overall performance of the multi-task system. As the weight of the sentiment task increases from 1.0 to 1.2, the model demonstrates steady improvement across all three evaluation metrics: aspect recognition F1 increases from 86.0% to 86.7%, sentiment classification accuracy increases from 88.3% to 89.3%, and Macro- F1 also increases from 85.2% to 85.9%. When the weight is further increased to 1.5 and 2.0, the sentiment classification accuracy still shows a slight improvement (reaching a maximum of 89.7%), but aspect recognition performance declines significantly (falling to 85.9% and 85.1%, respectively), causing Macro-F1 to drop to 85.3% and 84.7%, respectively. This indicates that overemphasizing the sentiment task weakens the learning effect of the aspect recognition subtask, thus disrupting the balance among the multi-task system. In summary, a weight of 1.2 for the sentiment task achieves the best trade-off between aspect recognition and sentiment classification, resulting in the highest overall performance of the model. Therefore, this paper ultimately adopts a weighting ratio of 1:1.2 for the joint loss.

V. CONCLUSION

This study constructs a domain-tuned RoBERTa model that accurately extracts fine-grained sentiment features related to library collections and services from reader feedback text, providing a technological breakthrough for library service evaluation and optimization. Experimental results show that the model achieves an F1 score of 86.7% on aspect recognition tasks and an accuracy of 89.3% for the joint classification of sentiment polarity and intensity, significantly outperforming mainstream baseline models such as BERT and TextCNN. Practically, this research provides data-driven decision support for collection resource optimization and service process improvement. By quantitatively analyzing reader sentiment regarding resource availability, service responsiveness, and spatial experience, this study identified “Electronic Resource Access” and “Book Acquisition Speed” as the two pain points with the highest concentration of negative sentiment, providing clear targets for service improvement. Based on the model’s feedback,

libraries can accurately identify service shortcomings and strengths, making targeted adjustments to collection structure, optimizing services, and improving resource allocation efficiency. The findings of this research indicate a shift in library management models, from traditional empirical judgment-based management to a refined operational model based on data insights, which supports advanced engineering disciplines by ensuring that library resources essential for technical research and professional education are precisely optimized.

AUTHOR CONTRIBUTIONS

Chenhui Song designed, collected and analyzed the data, and drafted the manuscript. Chenhui Song conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Chenhui Song participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] O. Onunka, T. Onunka, A. A. Fawole, I. J. Adeleke, and C. Daraojimbae, "Library and information services in the digital age: Opportunities and challenges," Acta Informatica Malaysia, vol. 7, no. 1, pp. 113-121, 2023, doi: 10.26480/aim.02.2023.113.121.

[2] S. J. Jain and P. K. Behera, "Visualizing the academic library of the future based on collections, spaces, technologies, and services," International Journal of Information Science & Management, vol. 21, no. 1, pp. 217-241, 2023, doi: 10.22034/ijism.2023.700794.

[3] S. Kaur and R. Chakravarty, "Analytics for measuring library use and satisfaction of mobile apps," Library Hi Tech News, vol. 38, no. 4, pp. 10-12, 2021, doi: 10.1108/LHTN-04-2021-0014.

[4] Y. A. Ajani, E. K. Adefila, S. A. Olarongbe, R. T. Enakrire, and N. Rabiu, "Big data and the management of libraries in the era of the fourth industrial revolution: Implications for policymakers," Digital Library Perspectives, vol. 40, no. 2, pp. 311-329, 2024, doi: 10.1108/DLP-10-2023-0083.

[5] K. K. Twum, M. Adams, S. Budu, and R. A. Anati Budu, "Achieving university libraries user loyalty through user satisfaction: The role of service quality," Journal of Marketing for Higher Education, vol. 32, no. 1, pp. 54-72, 2022, doi: 10.1080/08841241.2020.1825030.

[6] C. Mubofu, "Experiences, purposes, satisfaction and missing library services as predictors of library information services provision: A case study," Alexandria, vol. 33, no. 1, pp. 64-79, 2023, doi: 10.1177/095574902412447.

[7] R. Miao, "Intelligent library service transformation strategies based on convolutional networks in the information age in the context of digital transformation," Journal of Information & Knowledge Management, vol. 23, no. 2, Art. no. 2450028, 2024, doi: 10.1142/S021964922450028X.

[8] V. J. H. Reji, "Digital transformation: Libraries implementing advanced IT solutions like AI and data analytics to improve services," Academic Research Journal of Science and Technology, vol. 1, no. 7, pp. 1-13, 2025.

[9] L. Ma, M. Li, W. N. Zhang, J. Li, and T. Liu, "Unstructured text enhanced open-domain dialogue system: A systematic survey," ACM Transactions on Information Systems, vol. 40, no. 1, pp. 1-44, 2021, doi: 10.1145/3464377.

[10] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, no. 7, pp. 5731-5780, 2022, doi: 10.1007/s10462-022-10144-1.

[11] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, "Generalizing sentiment analysis: A review of progress, challenges, and emerging directions," Social Network Analysis and Mining, vol. 15, no. 1, pp. 1-28, 2025, doi: 10.1007/s13278-025-01461-8.

[12] M. Hu, H. Gao, Y. Wu, Z. Su, and S. Zhao, "Fine-grained domain adaptation for aspect category level sentiment analysis," IEEE Transactions on Affective Computing, vol. 14, no. 4, pp. 2839-2850, 2022, doi: 10.1109/TAFFC.2022.3228695.

[13] M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications and future scopes," Artificial Intelligence Review, vol. 56, no. 11, pp. 12505-12560, 2023, doi: 10.1007/s10462-023-10442-2.

[14] S. K. Bichha, S. P. Singh, A. Choudhary, and K. Sahani, "Sentiment analysis of customer feedback: A pathway to emotion-centric service optimization," Decision-Making, vol. 13, no. 14, pp. 15, 2025, [Online]. Available: https://www.jsiar.com/2025-April/JSIAR-A-25-04050.pdf.

[15] S. A. Umezurike, O. V. Akinrinoye, O. T. Kufile, A. Y. Onifade, B. O. Otokiti, and O. G. Ejike, "Cross-platform sentiment analytics for unified customer feedback in digital business environments," International Journal of Scientific Research in Science and Technology, vol. 12, no. 3, pp. 1224-1235, 2025, doi: 10.54660/.JFMR.2025.6.2.41-47.

[16] O. Chamorro-Atalaya, L. Sobrino-Chunga, R. Guerrero-Carranza, A. Vargas-Díaz, and C. Poma-Garcia, "Student satisfaction in the context of hybrid learning through sentiment analysis," International Journal of Evaluation and Research in Education, vol. 13, no. 2, pp. 831-841, 2024, doi: 10.11591/ijere.v13i2.26717.

[17] S. Sarkar, "A comparative study of different natural language processing techniques for sentiment analysis and opinion mining," Iconic Research and Engineering Journals, vol. 8, no. 12, pp. 1695-1708, 2025, [Online]. Available: https://www.irejournals.com/paper-details/1709305.

[18] W. Li and M. Zhang, "Optimizing reading comprehension and understanding in learning by reading systems through natural language processing and user feedback," Studies in Knowledge Discovery, Intelligent Systems, and Distributed Analytics, vol. 14, no. 11, pp. 26-39, 2024, [Online]. Available: https://edgescholar.com/index.php/SKDISDA/article/view/e-2024-11-10.

[19] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: An overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," Artificial Intelligence Review, vol. 55, no. 7, pp. 5543-5574, 2022, doi: 10.1007/s10462-021-10134-9.

[20] M. M. Pakpahan, M. H. Dar, and M. N. S. Hasibuan, "Performance evaluation of machine learning algorithms in aspect-based sentiment analysis on e-commerce user reviews," International Journal of Science, Technology & Management, vol. 6, no. 4, pp. 958-965, 2025, doi: 10.46729/ijstm.v6i4.1255.

[21] R. Abdullah and R. Sarno, "Aspect based sentiment analysis for explicit and implicit aspects in restaurant review using grammatical rules, hybrid approach, and SentiCircle," International Journal of Intelligent Engineering & Systems, vol. 14, no. 5, pp. 294-305, 2021, doi: 10.22266/ijies2021.1031.27.

[22] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 845-863, 2020, doi: 10.1109/TAFFC.2020.2970399.

[23] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-based sentiment analysis: A survey of deep learning methods," IEEE Transactions on Computational Social Systems, vol. 7, no. 6, pp. 1358-1375, 2020, doi: 10.1109/TCSS.2020.3033302.

[24] W. Liao, B. Zeng, X. Yin, and P. Wei, "An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa," Applied Intelligence, vol. 51, no. 6, pp. 3522-3533, 2021, doi: 10.1007/s10489-020-01964-1.

[25] J. Yan, P. Pu, and L. Jiang, "Emotion-RGC net: A novel approach for emotion recognition in social media using RoBERTa and graph neural networks," PLOS ONE, vol. 20, no. 3, Art. no. e0318524, 2025, doi: 10.1371/journal.pone.0318524.

[26] F. Alqarni, A. Sagheer, A. Alabbad, and H. Hamdoun, "Emotion-aware RoBERTa enhanced with emotion-specific attention and TF-IDF gating for fine-grained emotion recognition," Scientific Reports, vol. 15, no. 1, Art. no. 17617, 2025, doi: 10.1038/s41598-025-99515-6.
[27] H. Du, "Research on cloud storage biological data deduplication method based on Simhash algorithm," International Journal of Data Mining and Bioinformatics, vol. 27, no. 4, pp. 252-266, 2023, doi: 10.1504/IJDMB.2023.134296.
[28] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," ACM Transactions on Intelligent Systems and Technology, vol. 12, no. 5, pp. 1-32, 2021, doi: 10.1145/3465055.
[29] B. K. Nwobu and E. S. George, "Access to information resources using the online public access catalogue (OPAC) in the 21st century," The Catalyst Journal of Library and Information Literacy, vol. 3, no. 1, pp. 140–153, 2024, [Online]. Available: https://journals.journalsplace.org/index.php/CJLIL/article/view/539.
[30] J. Zeng, S. Cao, Y. Chen, P. Pan, and Y. Cai, "Measuring the interdisciplinary characteristics of Chinese research in library and information science based on knowledge elements," Aslib Journal of Information Management, vol. 75, no. 3, pp. 589-617, 2023, doi: 10.1108/AJIM-03-2022-0130.
[31] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, et al., "Domain-specific language model pretraining for biomedical natural language processing," ACM Transactions on Computing for Healthcare, vol. 3, no. 1, pp. 1-23, 2021, doi: 10.1145/3458754.
[32] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: A review," Artificial Intelligence Review, vol. 58, no. 6, pp. 1-49, 2025, doi: 10.1007/s10462-025-11162-5.
[33] J. Wang, J. X. Huang, X. Tu, J. Wang, A. J. Huang, M. T. R. Laskar, et al., "Utilizing BERT for information retrieval: Survey, applications, resources, and challenges," ACM Computing Surveys, vol. 56, no. 7, pp. 1-33, 2024, doi: 10.1145/3648471.
[34] V. K. Agbesi, W. Chen, S. B. Yussif, C. C. Ukwuoma, Y. H. Gu, and M. A. Al-antari, "MuTCELM: An optimal multi-TextCNNbased ensemble learning for text classification," Heliyon, vol. 10, no. 19, Art. no. e38515, 2024, doi: 10.1016/j.heliyon.2024.e38515.
[35] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, "MAPA BiLSTM-BERT: Multi-aspects position aware attention for aspect level sentiment analysis," Journal of Supercomputing, vol. 79, no. 10, Art. no. 11452, 2023, doi: 10.1007/s11227-023-05112-7.