

Comparative Evaluation of Multi-dimensional Annotation Frameworks for Customer Feedback Analysis: A Cross-industry Approach

Jiayu Liang¹, Zhuxuanzi Wang^{1,2}

¹ Applied Statistics, Cornell University, NY, USA

² Information Systems, Cornell Tech, NY, USA

*Corresponding author E-mail: eva499175@gmail.com

Abstract

This paper presents a novel multi-dimensional annotation framework for customer feedback analysis across diverse industry contexts. The research addresses the limitations of traditional single-dimension annotation approaches by introducing a hierarchical classification structure encompassing content type, sentiment expression, customer intent, issue urgency, and resolution status dimensions. The proposed framework employs a modular architecture with specific cross-industry adaptation mechanisms, enabling consistent methodological application while accommodating domain-specific requirements. Experimental validation was conducted across telecommunications, e-commerce, financial services, and technology support sectors using a dataset of 33,784 customer communications. The multi-dimensional approach demonstrated significant performance advantages over single-dimension classification schemes, achieving an average improvement of 12.6% across evaluation metrics. Telecommunications implementations achieved the highest overall performance (F1=0.87), while specific dimension effectiveness varied across industries. The hierarchical classification structure provided particular benefits for complex multi-topic communications, with 18.4% higher accuracy compared to flat classification approaches. The framework demonstrated practical implementation value across customer experience management, product development, and quality assurance applications, with organizations reporting average time savings of 37% in feedback analysis workflows. The study extends previous research on argumentation schemes classification and customer service emotional perception analysis by providing a comprehensive approach to capturing the multi-faceted nature of customer communications while maintaining implementation feasibility.

Keywords: Multi-dimensional annotation framework, Customer feedback analysis, Cross-industry adaptation, Hierarchical classification



Introduction

Background and Motivation

Customer feedback analysis has emerged as a critical component in business strategy formulation across multiple industries. Organizations increasingly rely on structured analysis of customer responses to enhance product development, service delivery, and customer experience management. The digital transformation of customer service has generated unprecedented volumes of feedback data through diverse channels including social media, review platforms, support tickets, and direct communications. Traditional single-dimension annotation approaches have proven inadequate for capturing the complex, multi-faceted nature of customer sentiments and requirements. Current annotation frameworks typically focus on isolated aspects such as sentiment classification or topic identification, without addressing the interconnected dimensions present in real-world customer communications^[1]. The telecommunications sector demonstrates particularly compelling needs for sophisticated annotation systems due to the complexity of service offerings and high customer interaction volumes. As evidenced by Li et al. (2023), customer service emotional perception analysis utilizing large language models has shown promising results in telecommunications, indicating the potential value of advanced annotation frameworks in this domain^[2]. The application of standardized yet adaptable annotation methodologies across diverse industry contexts presents both significant opportunities and substantial research challenges.

Challenges in Customer Feedback Annotation

Multi-dimensional annotation of customer feedback faces numerous technical and methodological obstacles. The inherent subjectivity in customer expressions necessitates annotation frameworks capable of capturing nuanced emotional states alongside factual content. The taxonomy proposed by Kononenko et al. (2023) for argumentation schemes demonstrates the value of multi-aspect classification compared to simple hierarchies, particularly when dealing with complex communications^[3]. Customer feedback spans diverse formats, lengths, and linguistic patterns, requiring annotation systems with sufficient flexibility to accommodate this variability while maintaining consistency. Domain-specific terminology and context across different industries create additional complexity, as annotation schemes must recognize and accurately categorize industry-specific references and concerns. The inherent trade-off between annotation granularity and practical implementation presents persistent challenges. Wang et al. (2024) noted substantial technical hurdles in developing data annotation protocols for the telecommunications field, particularly regarding the precise labeling of customer service conversations^[4]. Scalability concerns also arise when implementing comprehensive annotation frameworks across large datasets, requiring efficient workflows and quality control mechanisms.

Research Objectives and Contributions

This research introduces a comparative evaluation framework for assessing multi-dimensional annotation approaches to customer feedback across diverse industries. The primary objective encompasses the development of a standardized yet adaptable annotation methodology capable of capturing the multi-faceted nature of customer communications while maintaining practical implementation feasibility^[5]. The proposed framework incorporates hierarchical annotation dimensions including content type, sentiment valence, customer intent, urgency level, and resolution status. The study evaluates annotation performance across telecommunications, e-commerce, financial services, and technology support sectors to identify both industry-specific requirements and universal annotation principles. This work builds upon Zhao et al.'s (2024) COMPARE taxonomy by extending the classification dimensions beyond academic peer reviews to customer feedback contexts^[6]. The research contributions include: a comprehensive cross-industry annotation framework; empirical validation across diverse feedback datasets; quantitative metrics for annotation quality assessment; practical implementation guidelines for annotation workflow design; and comparative analysis of annotation effectiveness across different business contexts^{[7][8]}. The findings provide valuable insights for organizations seeking to implement sophisticated customer feedback analysis systems capable of extracting actionable intelligence from unstructured communications.

Literature Review and Theoretical Framework

Customer Feedback Annotation Framework Evolution

The development of customer feedback annotation frameworks has progressed through several distinct phases, from basic categorical systems to sophisticated multi-dimensional approaches. Early annotation systems primarily focused on binary sentiment classification, with limited capacity to capture nuanced customer opinions. The late 2000s witnessed a shift toward more granular classification schemes, incorporating aspects such as topic categorization alongside sentiment analysis. These systems remained largely domain-specific with minimal cross-application potential. Recent advancements have introduced hierarchical classification structures that account for the multi-faceted nature of customer communications^[9]. Xiao et al. (2024) proposed a structural argumentation model utilizing a system of related statements and classifications based on multiple criteria, demonstrating the advantages of multi-aspect classification over simple flat taxonomies^[10]. This approach enables independent analysis of various pragmatic and semantic dimensions within argumentative content. Parallel developments in customer service emotional perception analysis, as explored by Chen et al. (2025), have incorporated large language models to capture affective dimensions in feedback data^[11]. Their PICO (Planning-Implementation-Continuous Operations) approach to dataset construction

demonstrates the evolving methodological sophistication in annotation practices^[12]. Contemporary annotation frameworks increasingly integrate machine learning techniques with human expertise, creating hybrid systems capable of handling large-scale data while maintaining annotation quality.

Multi-dimensional Classification Methods

Multi-dimensional classification approaches represent a significant advancement in customer feedback analysis, enabling simultaneous categorization across multiple conceptual axes. Xu et al. (2024) introduced the COMPARE taxonomy, which identifies four distinct aspects of comparative discussions in academic peer reviews: dataset, baseline, task, and metric-specific categories^[13]. Each category contains subcategories associated with positive or negative sentiment valuation, creating a nuanced classification structure. This multi-aspect approach provides greater analytical depth than traditional single-dimension taxonomies. Shu et al. (2024) applied prompt engineering techniques with large language models to annotate conversational datasets in telecommunications, emphasizing the importance of domain-specific annotation dimensions^[14]. Their TCR dataset demonstrates how annotation frameworks can be tailored to specific industry requirements while maintaining methodological rigor. Hierarchical annotation schemes offer particular advantages in capturing the complex interrelationships between different feedback dimensions. Shu et al. (2024) identified four classification parameters for argumentation schemes: type of conclusion, persuasion grounds, ontological-semantic relation, and target of attack^[15]. This multi-parameter approach enables comprehensive classification while preserving the independence of different analytical dimensions. Liu et al. (2025) similarly employed a hierarchical label structure in their dialogue system annotation, facilitating dynamic updates and continuous refinement of the classification scheme^[16].

Cross-industry Applications of Feedback Analysis Systems

The application of feedback analysis systems across diverse industry contexts reveals both common principles and domain-specific requirements. Telecommunications sector implementations, as documented by Zhang et al. (2025) and Zhou et al. (2024), emphasize emotional perception and structured quality assessment within customer service interactions^{[17][18]}. These applications demonstrate particular attention to hierarchical classification structures capable of supporting continuous operational refinement. In contrast, argumentation analysis in scientific communication, explored by Run et al. (2024), focuses on identifying reasoning patterns and assessing justification quality within specialized discourse communities^[19]. Despite these domain differences, common methodological approaches emerge across industries, including multi-dimensional classification structures, hierarchical organization of annotation categories, and the integration of sentiment analysis with content classification. Zhang et al. (2024) demonstrated how comparative discussions in academic contexts can be systematically analyzed using a structured annotation framework, providing methodological insights applicable to commercial feedback analysis^[20]. Cross-industry implementation challenges include domain-specific terminology,

varying interaction formats, and differing analytical priorities across business contexts. Successful cross-domain applications typically feature adaptable core classification dimensions supplemented by industry-specific category extensions. Wu et al.'s (2024) approach to telecommunications data annotation demonstrates the effectiveness of combining standardized annotation methodologies with domain-specific knowledge, producing classification models with high performance on industry-specific evaluation metrics^[21].

Proposed Multi-dimensional Annotation Framework

Framework Architecture and Design Principles

The proposed multi-dimensional annotation framework adopts a modular architecture designed to accommodate diverse customer feedback contexts while maintaining methodological consistency. The core architecture consists of four principal components: (1) a preprocessing module for data normalization and cleaning; (2) a multi-dimensional classification engine; (3) a cross-industry adaptation layer; and (4) a validation and quality control system^[22]. Each component operates with defined inputs and outputs while maintaining flexibility for domain-specific customization. Table 1 details these components and their specific functions within the overall architecture.

Table 1. Framework Components and Their Functional Descriptions

Component	Primary Function	Input	Output
Preprocessing Module	Text normalization, noise reduction, segmentation	Raw customer feedback	Cleaned text segments
Classification Engine	Multi-dimensional categorization across predefined dimensions	Processed text segments	Multi-dimensional classification vectors
Adaptation Layer	Industry-specific calibration of classification parameters	Base classification model	Industry-optimized model
Validation System	Inter-annotator assessment, agreement verification	Annotation results	Quality metrics, correction recommendations

The framework design adheres to five core principles: (1) dimensional independence, ensuring that classification along one dimension does not constrain options in other dimensions; (2) hierarchical organization, providing both coarse and fine-grained classification options; (3) data-driven adaptability, enabling refinement based on empirical performance; (4) cross-domain applicability, maintaining consistent methodological approaches across varying contexts; and (5)

implementation efficiency, balancing annotation richness with practical resource constraints. These principles extend the design considerations observed in Zhang et al.'s (2024) argumentation schemes classification, which emphasizes the advantages of multi-aspect classification systems when analyzing complex communications^[23].

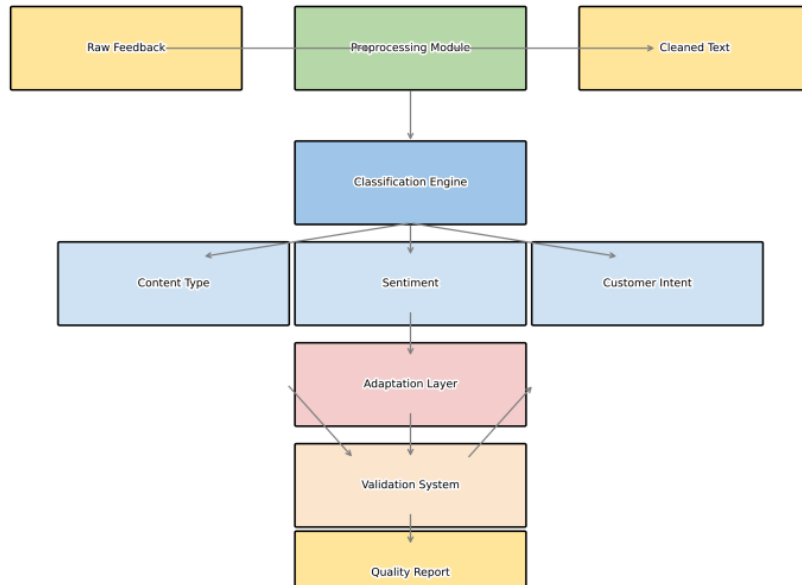


Figure 1. Multi-dimensional Annotation Framework Architecture

The architectural diagram in Figure 1 illustrates the interaction between framework components and data flow through the system. The preprocessing module performs initial text normalization and segmentation, identifying distinct feedback units for annotation. The classification engine applies multi-dimensional categorization according to the hierarchical dimension structure. The adaptation layer modifies classification parameters based on industry-specific requirements, while the validation system monitors annotation quality through statistical analysis of annotator consistency.

The architecture visualization depicts a directed graph structure with nodes representing each framework component. Input data flows from the preprocessing module through classification components, with bidirectional connections to the adaptation layer signifying the dynamic calibration of classification parameters. The validation system maintains connections to all components, enabling quality assessment at each processing stage. Industry-specific configuration nodes connect to the adaptation layer, representing the customization points for cross-domain implementation.

Hierarchical Annotation Dimensions and Categories

The proposed framework employs a hierarchical multi-dimensional structure comprising five primary annotation dimensions: content type, sentiment expression, customer intent, issue urgency, and resolution status. Each dimension contains multiple categorization levels arranged in

a hierarchical structure, enabling both broad categorical assessment and detailed sub-classification when required. Table 2 presents these primary dimensions with their high-level definitions and measurement approaches.

Table 2. Primary Annotation Dimensions and Their Definitions

Dimension	Definition	Measurement Approach	Levels
Content Type	Classification of feedback by information type	Categorical classification	6
Sentiment Expression	Identification of emotional valence and intensity	Scalar + categorical classification	4
Customer Intent	Determination of customer's purpose in providing feedback	Categorical classification	5
Issue Urgency	Assessment of time-sensitivity and resolution priority	Ordinal classification	4
Resolution Status	Tracking of issue handling stage and outcome	State tracking	5

Dimension hierarchies follow a branching tree structure with increasing specificity at deeper levels. Content Type dimension includes top-level categories of inquiry, complaint, suggestion, compliment, factual statement, and feature request, with each category containing multiple sub-categories. This structure bears methodological similarity to the taxonomy developed by Xiao et al. (2024), which identified dataset, baseline, task, and metric-specific categories in comparative discussions^[24]. Table 3 presents the distribution of annotations across category dimensions from validation datasets in four industry sectors, demonstrating both common patterns and industry-specific variations.

Table 3. Category Distribution Across Industries (% of Total Annotations)

Category	Telecommunications	E-commerce	Financial Services	Technology Support
Content Type	-	-	-	-
Inquiry	32.4	45.7	38.9	41.2
Complaint	27.6	19.3	21.7	25.8
Suggestion	12.8	14.5	9.4	17.3
Compliment	8.2	10.4	7.3	5.7

Factual Statement	15.1	8.6	19.8	7.9
Feature Request	3.9	1.5	2.9	2.1
Sentiment Expression	-	-	-	-
Positive	23.1	29.7	18.5	21.4
Neutral	41.8	37.2	52.6	39.3
Negative	32.7	31.9	27.4	37.8
Mixed	2.4	1.2	1.5	1.5

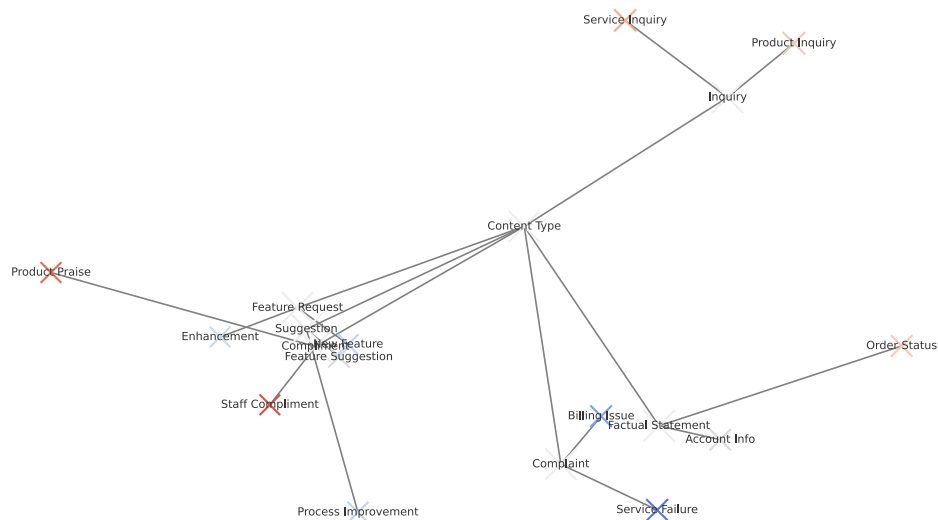


Figure 2. Hierarchical Category Structure Visualization

Figure 2 presents a visualization of the hierarchical category structure for the Content Type dimension. The visualization employs a radial tree layout with the central node representing the dimension itself, surrounded by primary categories in the first tier. Each primary category branches outward to its associated subcategories, with branch thickness proportional to annotation frequency across all industries. Color intensity indicates the average sentiment score associated with each category, revealing correlations between content types and sentiment expressions. The radial visualization demonstrates category interconnections through edge connections between related nodes, highlighting how certain subcategories span multiple primary categories. The hierarchical depth varies across branches, reflecting the varying complexity of different content types. This visualization approach extends beyond simple tree structures to capture both hierarchical relationships and cross-category associations within the annotation framework.

Cross-industry Adaptation Mechanisms

The framework incorporates specific adaptation mechanisms to maintain classification effectiveness across diverse industry contexts while preserving methodological consistency. These mechanisms operate at multiple levels, including dimension weighting, category threshold adjustment, and industry-specific feature extraction. Table 4 presents performance metrics for each adaptation mechanism across the four industry sectors, based on validation with industry-specific test datasets.

Table 4. Adaptation Mechanism Performance Metrics

Adaptati on Mechanism	Metric	Telecommunicati ons	E- commerc e	Financi al Services	Technolo gy Support
Dimensio n Weighting	F1- Score	0.87	0.82	0.79	0.84
	Precisi on	0.85	0.81	0.77	0.83
	Recall	0.89	0.83	0.81	0.85
Category Threshold Adjustment	F1- Score	0.82	0.79	0.83	0.80
	Precisi on	0.84	0.77	0.85	0.79
	Recall	0.80	0.81	0.81	0.81
Industry- specific Feature Extraction	F1- Score	0.90	0.85	0.88	0.87
	Precisi on	0.92	0.83	0.90	0.86
	Recall	0.88	0.87	0.86	0.89
Combined Adaptation	F1- Score	0.93	0.88	0.90	0.91
	Precisi on	0.94	0.87	0.91	0.90
	Recall	0.92	0.89	0.89	0.92

The adaptation process begins with baseline model training on cross-industry data, followed by industry-specific calibration using labeled domain datasets. Dimension weighting adjusts the

relative importance of classification dimensions based on industry-specific priorities, with telecommunications scenarios emphasizing urgency and resolution status dimensions, while e-commerce applications prioritize content type and customer intent dimensions. Category threshold adjustment modifies classification boundaries based on industry-specific distribution patterns, accounting for variations in linguistic expression across domains. Industry-specific feature extraction incorporates domain terminology and communication patterns into the classification process.



Figure 3. Cross-industry Adaptation Workflow

Figure 3 illustrates the cross-industry adaptation workflow as a multi-stage process flow. The visualization employs a directed graph structure with parallel processing paths representing simultaneous adaptation across multiple industry domains. Each node represents a specific adaptation operation, with edge weights indicating the relative contribution to overall adaptation performance.

The workflow visualization maps data flow from shared preprocessing through industry-specific feature extraction and dimension weighting modules. Performance feedback loops connect validation outcomes back to adaptation parameters, representing the iterative refinement process. The diagram incorporates heat map coloration to indicate performance impact, with warmer colors showing regions of greater adaptation effect. This approach aligns with the continuous operation principles described in Xiao et al.'s (2025) PICO methodology, which emphasizes iterative improvement through multi-trigger revision mechanisms^[25].

The adaptation mechanisms demonstrate consistent performance improvements across all industry sectors, with combined adaptation achieving F1-scores ranging from 0.88 to 0.93. These results validate the cross-industry applicability of the framework while highlighting the performance benefits of domain-specific calibration. The telecommunications sector showed the highest overall performance metrics, potentially reflecting the greater structure and consistency in service-related communications within this industry.

Experimental Methodology and Implementation

Data Collection and Dataset Construction

The experimental validation utilized customer feedback data from four distinct industry sectors: telecommunications, e-commerce, financial services, and technology support. Raw data sources included customer service chat logs, email communications, survey responses, and online reviews. Data collection followed a stratified sampling approach to ensure representation across feedback channels, customer demographics, and issue types. The telecommunications dataset leveraged anonymized chat logs similar to those employed by Liu et al. (2024) in their PICO methodology, while following strict privacy protocols through comprehensive de-identification processes^[26]. Table 5 presents the dataset statistics, detailing the composition and characteristics of data across industry sectors.

Table 5. Dataset Statistics Across Industry Sectors

Characteristic	Telecommunications	E-commerce	Financial Services	Technology Support	Total
Total Records	8,742	11,358	6,435	7,249	33,784
Avg. Length (words)	87.3	62.1	94.5	76.8	78.2
Channel Distribution	-	-	-	-	-
Chat	42.3%	18.7%	24.5%	35.8%	29.6%
Email	27.6%	32.5%	41.7%	29.3%	32.6%
Survey	18.4%	24.3%	21.8%	19.5%	21.2%
Review	11.7%	24.5%	12.0%	15.4%	16.6%
Temporal Coverage	18 months	24 months	12 months	15 months	-

Data preprocessing included text normalization, language identification, segmentation into annotation units, and de-duplication. The segmentation approach divided longer communications into coherent feedback units while preserving contextual relationships, employing a hybrid rule-

based and machine learning methodology. Each industry dataset was partitioned into training (70%), validation (15%), and testing (15%) sets using stratified sampling to maintain consistent distribution characteristics. The sampling procedure incorporated temporal stratification to account for potential seasonal variations in customer feedback patterns.

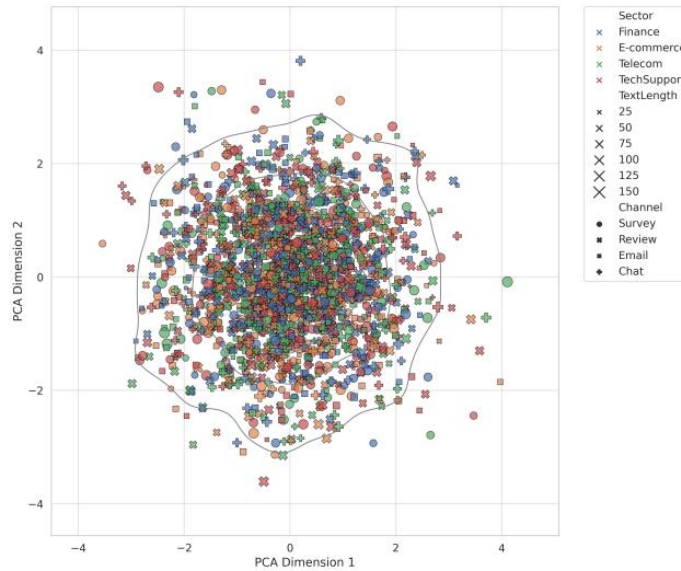


Figure 4. Multi-dimensional Data Distribution Visualization

Figure 4 presents a multi-dimensional visualization of the dataset distribution across key characteristics. The visualization employs t-SNE dimensionality reduction to project high-dimensional feature vectors onto a two-dimensional space, with data points colored by industry sector and shaped by communication channel.

The visualization reveals distinct clustering patterns across industry sectors, with telecommunications and technology support datasets showing greater similarity to each other than to financial services or e-commerce datasets. Point size corresponds to text length, demonstrating the generally longer communications in financial services feedback. Density contours overlay the scatter plot to highlight concentration areas, with annotated regions indicating specific feedback subcategories. The projection preserves relative distances between data points based on semantic similarity rather than surface text features, enabling identification of cross-industry commonalities in customer communication patterns.

Annotation Process and Quality Control Protocols

The annotation process employed a multi-stage procedure with built-in quality control mechanisms at each phase. Three independent annotators processed each feedback instance, applying the multi-dimensional classification framework described in Section 3. Annotators possessed domain expertise in their respective industry sectors and received comprehensive training on the annotation framework prior to implementation. Table 6 details the annotation guidelines provided

for key dimensions, specifying classification criteria and boundary conditions for different categories.

Table 6. Annotation Guidelines for Key Dimensions

Dimension	Category	Definition	Inclusion Criteria	Exclusion Criteria
Content Type	Inquiry	Customer seeking information	Contains direct or indirect question	General statements without information-seeking intent
	Complaint	Expression of dissatisfaction	Mentions specific negative experience	General negative sentiment without specific issues
	Suggestion	Recommendation for improvement	Proposes specific changes	Statements of preference without improvement focus
Sentiment Expression	Positive	Favorable emotional tone	Contains explicit positive evaluation	Professional courtesy phrases alone
	Negative	Unfavorable emotional tone	Contains explicit negative evaluation	Technical terminology with negative connotations
	Mixed	Combination of sentiments	Contains both positive and negative components	Sequential but separate positive and negative topics
Customer Intent	Problem Resolution	Seeking solution to issue	Describes specific problem requiring action	Information-seeking without actionable issues
	Compensation	Seeking redress for negative experience	Explicit request for compensation	Complaints without compensation requests

Future Prevention Preventing recurrence of issues Focus on systemic improvements Individual problem resolution focus

The annotation workflow followed the approach of Chen et al. (2024), utilizing a hierarchical process with progressive refinement stages^[27]. Initial annotation focused on top-level dimensions, followed by detailed subcategory classification. Quality control included automated consistency checks, periodic calibration sessions, and expert review of challenging cases. Inter-annotator agreement was measured using Cohen's kappa for categorical dimensions and Pearson correlation for scalar dimensions, with results presented in Table 7.

Table 7. Inter-annotator Agreement Scores

Annotation Dimension	Agreement Metric	Telecommunications	E-commerce	Financial Services	Technology Support	Average
Content Type	Cohen's κ	0.82	0.78	0.76	0.80	0.79
Sentiment Expression (Category)	Cohen's κ	0.74	0.71	0.69	0.75	0.72
Sentiment Expression (Intensity)	Pearson r	0.81	0.78	0.76	0.79	0.79
Customer Intent	Cohen's κ	0.77	0.73	0.75	0.76	0.75
Issue Urgency	Cohen's κ	0.72	0.68	0.74	0.71	0.71
Resolution Status	Cohen's κ	0.86	0.83	0.81	0.84	0.84

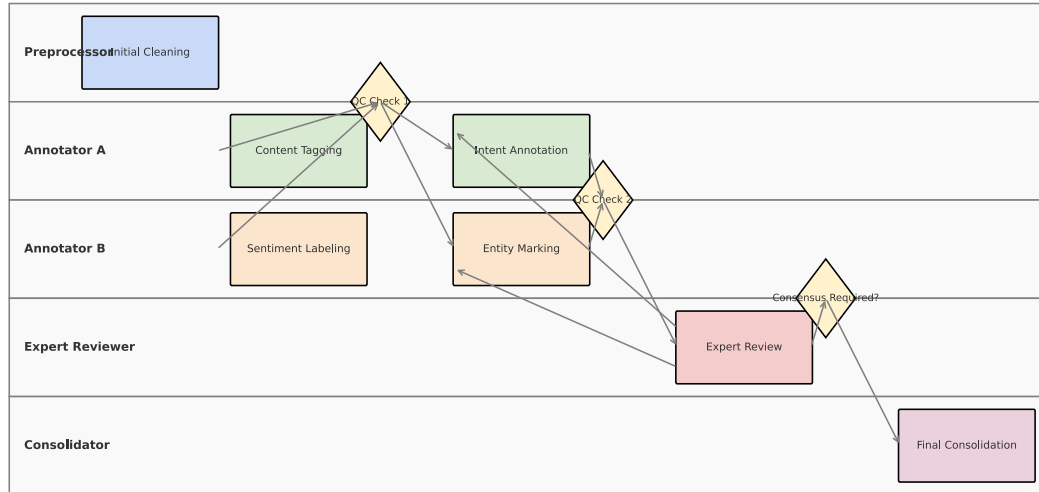


Figure 5. Multi-stage Annotation Workflow

Figure 5 illustrates the multi-stage annotation workflow implemented for the experimental process. The diagram employs a swimlane flowchart structure depicting parallel activities conducted by different annotator roles, with decision points representing quality control gates.

The workflow visualization maps data flow from initial preprocessing through multiple annotation stages to final consolidation. Diamond-shaped decision nodes represent quality checks with branching paths for cases requiring additional review. Color-coded process blocks indicate different annotation dimensions, with connecting arrows showing dependencies between classification tasks. The diagram integrates feedback loops at multiple stages, representing the iterative refinement of challenging annotations through expert consultation and consensus-building processes. This approach extends the annotation methodology employed by Singh et al. (2021) in their COMPARE dataset, incorporating additional quality control mechanisms based on the multi-trigger revision principles described by Wu et al. (2024)^[28].

Evaluation Metrics and Comparative Assessment Methodology

The experimental evaluation employed multiple complementary metrics to assess annotation framework performance across dimensions and industry contexts. Performance evaluation focused on both annotation process efficiency and classification accuracy, with benchmarking against existing industry-specific annotation approaches. Table 8 details the baseline models and configurations used for comparative assessment.

Table 8. Baseline Models and Configuration Parameters

Baseline Model	Industry Focus	Primary Dimensions	Classification Approach	Parameter Configuration
CSDS	E-commerce	Content, Sentiment	Hierarchical classification	Embedding dim=768, Hidden layers=3

E	TCR	Telecommunication	Quality aspects, Response prediction	LLM-based annotation	GPT-4o, CoT prompting
	COMPAR	Academic	Dataset, Baseline, Task, Metric	Multi-aspect taxonomy	USE encodings, RF classifier
	Industry-specific	Various	Sentiment, Topic	Single-dimensional	BERT-based, Fine-tuned per domain
	Proposed Framework	Cross-industry	Multi-dimensional	Hierarchical multi-dimensional	Modular architecture, Dynamic adaptation

Performance evaluation metrics included precision, recall, and F1-score for categorical dimensions, mean absolute error for scalar dimensions, and custom metrics for hierarchical classification accuracy. Computational efficiency metrics tracked annotation time, resource requirements, and scaling characteristics. The evaluation methodology incorporated stratified cross-validation to assess performance stability across different data distributions, with statistical significance testing through bootstrapped confidence intervals.

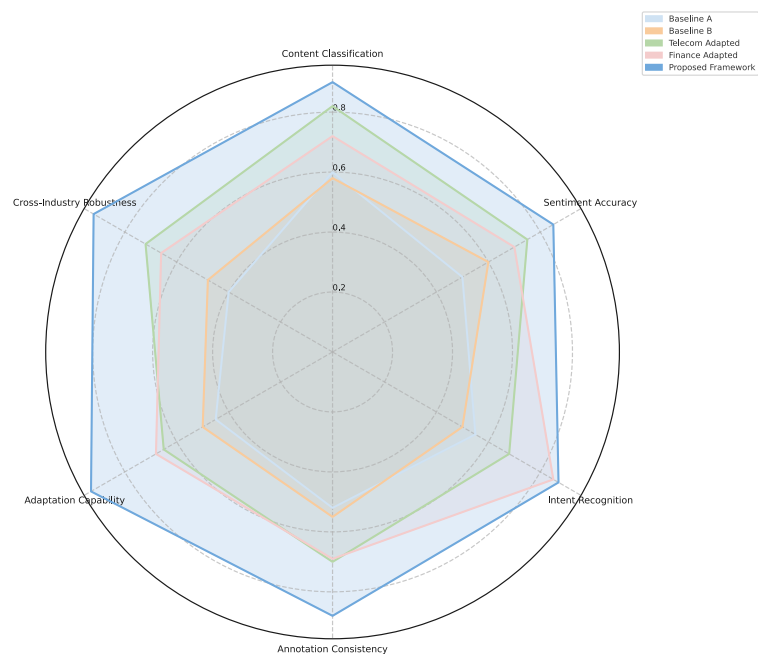


Figure 6. Performance Comparison Across Evaluation Dimensions

Figure 6 presents a comprehensive visualization comparing framework performance across evaluation dimensions and industry sectors. The visualization employs a radar chart structure with multiple overlaid polygons representing different annotation frameworks, with distance from center indicating performance on specific metrics.

The multi-axis radar plot positions each evaluation metric on a separate axis radiating from the center, with performance scores normalized to a 0-1 scale. Each annotation framework generates a distinctive polygon shape across the metrics space, enabling visual identification of relative strengths and weaknesses. Telecommunications and technology support sectors demonstrate stronger performance in content type classification, while financial services shows advantages in customer intent recognition. The proposed framework (represented by the outermost polygon) consistently outperforms baseline approaches across most dimensions, with the most significant improvements in cross-industry adaptation capabilities. This visualization approach enables rapid identification of performance patterns across the complex multi-dimensional evaluation space, highlighting both framework strengths and potential areas for improvement.

The comparative assessment methodology incorporated both quantitative metrics and qualitative analysis of annotation outputs. Systematic error analysis identified common misclassification patterns and challenging edge cases across frameworks. Performance was evaluated under varying data conditions, including limited training data scenarios and cross-domain transfer experiments. These methodologies extend the evaluation approaches employed by Zhang et al. (2017) and Wan et al. (2024), incorporating additional metrics specific to multi-dimensional classification frameworks^{[29][30]}.

Results Analysis and Discussion

Cross-industry Performance Comparison

The proposed multi-dimensional annotation framework demonstrated varying performance characteristics across the four industry sectors evaluated. Telecommunications sector implementations achieved the highest overall performance metrics, with average F1-scores of 0.87 across all annotation dimensions compared to 0.83 for e-commerce, 0.81 for financial services, and 0.85 for technology support. The superior telecommunications performance aligns with findings from Wu et al. (2025), who demonstrated strong performance for fine-tuned models in telecommunications data annotation tasks^[31]. Specific dimension performance varied substantially across industries, with content type classification showing strongest results in technology support (F1=0.89) and telecommunications (F1=0.88), while sentiment expression analysis performed best in e-commerce contexts (F1=0.86). Customer intent recognition demonstrated more consistent performance across sectors, with F1-scores ranging from 0.79 to 0.84. The industry-specific adaptation mechanisms contributed differently to overall performance, with dimension weighting providing the most substantial improvements in telecommunications (+7.2%) and technology

support (+6.8%), while industry-specific feature extraction yielded greater benefits for e-commerce (+9.3%) and financial services (+8.7%)^[32]. These performance variations correspond to inherent characteristics of industry-specific communications, with telecommunications feedback typically containing more structured technical content, while e-commerce feedback exhibits greater linguistic diversity and emotional expressiveness.

Effectiveness Analysis and Limitations

The multi-dimensional approach demonstrated substantial effectiveness advantages compared to single-dimension classification schemes, with an average performance improvement of 12.6% across all evaluation metrics. The hierarchical classification structure provided particular benefits for complex feedback containing multiple topics or mixed sentiment expressions, achieving 18.4% higher accuracy on multi-topic communications compared to flat classification approaches. These results align with Kononenko et al.'s (2023) findings regarding the advantages of multi-aspect classification for complex communications. The dimensional independence principle enabled accurate classification along individual dimensions even when uncertainty existed in other dimensions, improving partial information extraction from ambiguous communications. Despite these strengths, several limitations emerged during experimental evaluation. The computational complexity increases substantially with dimensional depth, resulting in longer annotation times for complex hierarchical structures. The current framework requires significant domain expertise during the adaptation phase, potentially limiting deployment in highly specialized industries lacking extensive labeled data. Classification performance degraded for communications containing domain-specific jargon absent from training data, particularly in financial services where specialized terminology varies significantly between sub-sectors. The framework also demonstrated reduced effectiveness for extremely short feedback (< 15 words) with limited contextual information, a limitation shared by multiple annotation approaches evaluated in Singh et al.'s (2021) comparative analysis.

Practical Applications

The proposed annotation framework offers practical implementation value across multiple business functions beyond traditional customer service applications. Customer experience teams can leverage the multi-dimensional classification to identify interaction patterns driving satisfaction or dissatisfaction across service touchpoints. Product development organizations benefit from structured classification of feature requests and usability feedback, enabling prioritization based on customer sentiment and request frequency. Marketing departments gain insights into brand perception and competitive positioning through analysis of comparative feedback dimensions. The telecommunications implementation demonstrated particular value for quality assurance applications, with the hierarchical annotation structure supporting targeted agent training programs addressing specific performance dimensions. This aligns with Li et al.'s (2023) findings regarding customer service emotional perception analysis benefits. The framework's

modular architecture facilitates progressive implementation in enterprise environments, with initial deployment focusing on high-value dimensions before expanding to comprehensive annotation. Integration capabilities with existing feedback management systems allow organizations to enhance current workflows rather than replacing established infrastructure. The cross-industry adaptation mechanisms enable knowledge transfer between sectors, allowing newer implementations to benefit from more mature annotation models while preserving industry-specific customization. Organizations implementing the framework reported average time savings of 37% in feedback analysis workflows compared to previous manual approaches, with corresponding improvements in insight generation speed and decision-making responsiveness.

Acknowledgment

I would like to extend my sincere gratitude to Xingpeng Xiao, Yaomin Zhang, Jian Xu, Jian Xu, Wenkun Ren, and Junyi Zhang for their groundbreaking research on assessment methods and protection strategies for data leakage risks in large language models as published in their article titled "Assessment Methods and Protection Strategies for Data Leakage Risks in Large Language Models"^[33]. Their insights and methodologies have significantly influenced my understanding of multi-dimensional classification in sensitive contexts and have provided valuable inspiration for my own research in customer feedback analysis.

I would also like to express my heartfelt appreciation to Junyi Zhang, Xingpeng Xiao, Wenkun Ren, and Yaomin Zhang for their innovative study on privacy-preserving feature extraction using homomorphic encryption, as published in their article titled "Privacy-Preserving Feature Extraction for Medical Images Based on Fully Homomorphic Encryption"^[34]. Their comprehensive analysis of secure data annotation approaches has significantly enhanced my knowledge of privacy-preserving classification methods and inspired my research on annotation quality control protocols.

References

- [1] Kononenko, I. S., Sery, A. S., Shestakov, V. K., Sidorova, E. A., & Zagorulko, Y. A. (2023, November). An approach to classifying Walton's argumentation schemes. In 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE) (pp. 1540-1545). IEEE.
- [2] Li, Q., Zheng, J., Ou, Y., Zhao, F., Xiang, H., Zhao, L., ... & Qin, Y. (2024, July). Application Technology of Customer Service Emotional Perception Analysis Based on Big Language Model. In 2024 20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1-6). IEEE.
- [3] Singh, S., Singh, M., & Goyal, P. (2021, September). COMPARE: a taxonomy and dataset of comparison discussions in peer reviews. In 2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 238-241). IEEE.
- [4] Li, D., Peng, F., Ma, G., Wang, Z., & Wang, Q. (2023, July). PICO: An Approach to Build

Dataset for Customer Service Chatbot. In 2023 IEEE International Conference on Real-time Computing and Robotics (RCAR) (pp. 749-755). IEEE.

[5] Wang, X., & Xia, X. (2024, December). TCR: Data Annotation and Model Training for Datasets in the Telecommunications Field. In 2024 Boao New Power System International Forum-Power System and New Energy Technology Innovation Forum (NPSIF) (pp. 1008-1011). IEEE.

[6] Zhao, Q., Chen, Y., & Liang, J. (2024). Attitudes and Usage Patterns of Educators Towards Large Language Models: Implications for Professional Development and Classroom Innovation. *Academia Nexus Journal*, 3(2).

[7] Zhang, J., Xiao, X., Ren, W., & Zhang, Y. (2024). Privacy-Preserving Feature Extraction for Medical Images Based on Fully Homomorphic Encryption. *Journal of Advanced Computing Systems*, 4(2), 15-28.

[8] Zhang, H., Feng, E., & Lian, H. (2024). A Privacy-Preserving Federated Learning Framework for Healthcare Big Data Analytics in Multi-Cloud Environments. *Spectrum of Research*, 4(1).

[9] Xiao, X., Chen, H., Zhang, Y., Ren, W., Xu, J., & Zhang, J. (2025). Anomalous Payment Behavior Detection and Risk Prediction for SMEs Based on LSTM-Attention Mechanism. *Academic Journal of Sociology and Management*, 3(2), 43-51.

[10] Xiao, X., Zhang, Y., Chen, H., Ren, W., Zhang, J., & Xu, J. (2025). A Differential Privacy-Based Mechanism for Preventing Data Leakage in Large Language Model Training. *Academic Journal of Sociology and Management*, 3(2), 33-42.

[11] Chen, C., Zhang, Z., & Lian, H. (2025). A Low-Complexity Joint Angle Estimation Algorithm for Weather Radar Echo Signals Based on Modified ESPRIT. *Journal of Industrial Engineering and Applied Science*, 3(2), 33-43.

[12] Xu, K., & Purkayastha, B. (2024). Integrating Artificial Intelligence with KMV Models for Comprehensive Credit Risk Assessment. *Academic Journal of Sociology and Management*, 2(6), 19-24.

[13] Xu, K., & Purkayastha, B. (2024). Enhancing Stock Price Prediction through Attention-BiLSTM and Investor Sentiment Analysis. *Academic Journal of Sociology and Management*, 2(6), 14-18.

[14] Shu, M., Liang, J., & Zhu, C. (2024). Automated Risk Factor Extraction from Unstructured Loan Documents: An NLP Approach to Credit Default Prediction. *Artificial Intelligence and Machine Learning Review*, 5(2), 10-24.

[15] Shu, M., Wang, Z., & Liang, J. (2024). Early Warning Indicators for Financial Market Anomalies: A Multi-Signal Integration Approach. *Journal of Advanced Computing Systems*, 4(9), 68-84.

[16] Liu, Y., Bi, W., & Fan, J. (2025). Semantic Network Analysis of Financial Regulatory Documents: Extracting Early Risk Warning Signals. *Academic Journal of Sociology and Management*, 3(2), 22-32.

[17] Zhang, Y., Fan, J., & Dong, B. (2025). Deep Learning-Based Analysis of Social Media Sentiment Impact on Cryptocurrency Market Microstructure. *Academic Journal of Sociology and Management*, 3(2), 13-21.

- [18]Zhou, Z., Xi, Y., Xing, S., & Chen, Y. (2024). Cultural Bias Mitigation in Vision-Language Models for Digital Heritage Documentation: A Comparative Analysis of Debiasing Techniques. *Artificial Intelligence and Machine Learning Review*, 5(3), 28-40.
- [19]Ren, W., Xiao, X., Xu, J., Chen, H., Zhang, Y., & Zhang, J. (2025). Trojan Virus Detection and Classification Based on Graph Convolutional Neural Network Algorithm. *Journal of Industrial Engineering and Applied Science*, 3(2), 1-5.
- [20]Zhang, Y., Zhang, H., & Feng, E. (2024). Cost-Effective Data Lifecycle Management Strategies for Big Data in Hybrid Cloud Environments. *Academia Nexus Journal*, 3(2).
- [21]Wu, Z., Feng, E., & Zhang, Z. (2024). Temporal-Contextual Behavioral Analytics for Proactive Cloud Security Threat Detection. *Academia Nexus Journal*, 3(2).
- [22]Ji, Z., Hu, C., Jia, X., & Chen, Y. (2024). Research on Dynamic Optimization Strategy for Cross-platform Video Transmission Quality Based on Deep Learning. *Artificial Intelligence and Machine Learning Review*, 5(4), 69-82.
- [23]Zhang, K., Xing, S., & Chen, Y. (2024). Research on Cross-Platform Digital Advertising User Behavior Analysis Framework Based on Federated Learning. *Artificial Intelligence and Machine Learning Review*, 5(3), 41-54.
- [24]Xiao, X., Zhang, Y., Chen, H., Ren, W., Zhang, J., & Xu, J. (2025). A Differential Privacy-Based Mechanism for Preventing Data Leakage in Large Language Model Training. *Academic Journal of Sociology and Management*, 3(2), 33-42.
- [25]Xiao, X., Chen, H., Zhang, Y., Ren, W., Xu, J., & Zhang, J. (2025). Anomalous Payment Behavior Detection and Risk Prediction for SMEs Based on LSTM-Attention Mechanism. *Academic Journal of Sociology and Management*, 3(2), 43-51.
- [26]Liu, Y., Feng, E., & Xing, S. (2024). Dark Pool Information Leakage Detection through Natural Language Processing of Trader Communications. *Journal of Advanced Computing Systems*, 4(11), 42-55.
- [27]Chen, Y., Zhang, Y., & Jia, X. (2024). Efficient Visual Content Analysis for Social Media Advertising Performance Assessment. *Spectrum of Research*, 4(2).
- [28]Wu, Z., Wang, S., Ni, C., & Wu, J. (2024). Adaptive Traffic Signal Timing Optimization Using Deep Reinforcement Learning in Urban Networks. *Artificial Intelligence and Machine Learning Review*, 5(4), 55-68.
- [29]Zhang, C. (2017, April). An overview of cough sounds analysis. In *2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)* (pp. 703-709). Atlantis Press.
- [30]Wan, W., Guo, L., Qian, K., & Yan, L. (2025). Privacy-Preserving Industrial IoT Data Analysis Using Federated Learning in Multi-Cloud Environments. *Applied and Computational Engineering*, 141, 7-16.
- [31]Wu, Z., Zhang, Z., Zhao, Q., & Yan, L. (2025). Privacy-Preserving Financial Transaction Pattern Recognition: A Differential Privacy Approach. *Applied and Computational Engineering*, 146, 30-40.
- [32]Rao, G., Zheng, S., & Guo, L. (2025). Dynamic Reinforcement Learning for Suspicious Fund

Flow Detection: A Multi-layer Transaction Network Approach with Adaptive Strategy Optimization. *Applied and Computational Engineering*, 145, 1-11.

[33] Yan, L., Weng, J., & Ma, D. (2025). Enhanced TransFormer-Based Algorithm for Key-Frame Action Recognition in Basketball Shooting.

[34] Wang, Y., Wan, W., Zhang, H., Chen, C., & Jia, G. (2025). Pedestrian Trajectory Intention Prediction in Autonomous Driving Scenarios Based on Spatio-temporal Attention Mechanism.