electronics

MDPI

Article

A Hybrid Intention Recognition Framework with Semantic Inference for Financial Customer Service

Nian Cai [1], Shishan Li [1], Jiajie Xu [1], Yinfeng Tian [1], Yinghong Zhou [1,*] and Jiacheng Liao [2,*]

[1] School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China; cainian@gdut.edu.cn (N.C.); 2112303163@mail2.gdut.edu.cn (S.L.); 3122002030@mail2.gdut.edu.cn (J.X.); 2112103051@mail2.gdut.edu.cn (Y.T.)

[2] Foshan Talkiin Technology Co., Ltd., 5th Floor, Building 1, Wanbang Commercial Plaza, Daliang Street, Shunde District, Foshan 528300, China

* Correspondence: zhouyh@gdut.edu.cn (Y.Z.); liaojiacheng@talkiin.cn (J.L.)

Abstract: Automatic intention recognition in financial service scenarios faces challenges such as limited corpus size, high colloquialism, and ambiguous intentions. This paper proposes a hybrid intention recognition framework for financial customer service, which involves semi-supervised learning data augmentation, label semantic inference, and text classification. A semi-supervised learning method is designed to augment the limited corpus data obtained from the Chinese financial service scenario, which combines back-translation with BERT models. Then, a K-means-based semantic inference method is introduced to extract label semantic information from categorized corpus data, serving as constraints for subsequent text classification. Finally, a BERT-based text classification network is designed to recognize the intentions in financial customer service, involving a multi-level feature fusion for corpus information and label semantic information. During the multi-level feature fusion, a shallow-to-deep (StD) mechanism is designed to alleviate feature collapse. To validate our hybrid framework, 2977 corpus texts about loan service are provided by a financial company in China. Experimental results demonstrate that our hybrid framework outperforms existing deep learning methods in financial customer service intention recognition, achieving an accuracy of 89.06%, precision of 90.27%, recall of 90.40%, and an F1 score of 90.07%. This study demonstrates the potential of the hybrid framework to automatic intention recognition in financial customer service, which is beneficial for the improvement of the financial service quality.

Keywords: intention recognition; financial customer service; semi-supervised learning; label semantic inference; BERT

Academic Editor: Domenico Rosaci

Received: 6 January 2025
Revised: 22 January 2025
Accepted: 23 January 2025
Published: 25 January 2025

Citation: Cai, N.; Li, S.; Xu, J.; Tian, Y.; Zhou, Y.; Liao, J. A Hybrid Intention Recognition Framework with Semantic Inference for Financial Customer Service. *Electronics* **2025**, *14*, 495. https://doi.org/10.3390/electronics14030495

1. Introduction

The Internet of Things (IoT) provides a new generation of pervasive technologies and smart environments around the perimeter [1], and it has been widely applied in financial service scenarios. In the financial service scenarios, the IoT can be used to collect and analyze data to generate meaningful information to meet specific customer needs, which supports the digital bridge between financial institutions and their human customers. When the dialogue occurs between the financial customer service system and the customer, the service system can acquire the corpus data through the IoT, which can be intelligently analyzed to accurately recognize the customer's intention for subsequent responses [2]. For example, the phrase "I want to check my balance" in a dialogue can be understood by the system based on keywords such as "balance" and the context to accurately determine the

customer's intention. This accurate and timely customer service is significant for business and customer loyalty [3].

In recent years, automatic intention recognition based on natural language processing (NLP) has been applied in financial customer service to alleviate the heavy labor burden and improve the service experience and quality. Since corpus data in financial customer service are often scarce, data augmentation is usually used to improve model performance [4–7], involving back-translation, random deletion, and semantic substitution. Moreover, due to the high repetition and colloquial nature of corpus data in financial customer service, some inherent issues have affected intention recognition through customer service robots, such as short texts, inconsistent expressions with the same meaning (e.g., "check balance" and "I want to know my balance"), and ambiguous phrasing [8–11]. Therefore, accurate intention recognition remains a challenge in the field of natural language processing for financial customer service.

Traditional intention recognition methods commonly utilize the similarity measure of keywords or semantics, such as cosine similarity, the Jaccard coefficient, Pearson correlation coefficient, and Levenshtein distance, to automatically recognize customers' intentions [10]. However, some issues limit traditional intention recognition in many applications. Firstly, it is weak in contextual understanding, resulting in situations where the same keyword may have different meanings in different contexts, or different keywords may have the same meaning in different contexts. Secondly, it hardly captures the punctuations in the text and the tones in the voice dialogue, indicating its poor ability of handling ambiguity [11]. In contrast, due to its excellent abilities of self-learning and high adaptivity, deep learning has been successfully introduced in intention recognition [12]. Especially, due to its self-attention mechanism, Transformer can well capture long-term dependencies between words to understand the context [13–19], which is beneficial for the handling of ambiguity. As a Transformer-based model, bidirectional encoder representation from transformers (BERT) [14] is a promising approach to utilize the transformer model for intention recognition, in which pre-trained models are fine-tuned during training. However, in financial customer service, the corpus data are commonly not enough to train a reliable network for intention recognition. Also, extensive short texts carry less information and possibly make the network learn a number of redundant features. Furthermore, serious colloquial expressions combined with short texts most probably cause ambiguous semantics for the corpus data, which is not beneficial for intention recognition.

To address these issues, a hybrid intention recognition framework is designed to identify customer intentions in financial customer service, involving data augmentation, semantic inference, and text classification. To address the problem of insufficient corpus data in the dataset, a semi-supervised learning method is proposed to augment the corpus data. Specifically, the back-translation method is applied to achieve a large amount of low-quality augmented data from the original corpus data. Then, a BERT model is well trained on the original corpus data to predict labels for the back-translated augmented data. The corpus data with correctly predicted labels are finally mixed with the original corpus data to achieve high-quality augmented corpus data. To reduce the impact of severe colloquialism and short texts, a K-means-based semantic inference module is proposed to automatically extract the semantic information from corpus samples. Finally, a text classification network with two information branches is designed, which integrates corpus information and label semantic information via a specifically designed multi-level feature fusion to recognize customer intentions. An StD mechanism is incorporated to alleviate feature collapse.

In general, the contributions of our work are summarized as follows:

(1) To further improve intention recognition performance in financial service scenarios, a hybrid machine learning framework is designed for automatic customer intent recognition. The framework includes semi-supervised learning data augmentation, K-means-based label semantic inference, and a dual-branch text classification network. By combining traditional clustering methods with deep learning, K-means is used to find the approximately central text that represents the label semantics, which is then input into the text classification network, using the multi-level feature fusion approach that enhances the accuracy of intent recognition.
(2) A K-means-based label semantic inference scheme is utilized to integrate the knowledge related to label semantics into the feature information of the corpus texts by identifying the approximately central text of each semantic category, thereby alleviating the ambiguous semantics caused by the combination of colloquialism and short texts.
(3) A dual-stream text classification network is designed for intention recognition, incorporating corpus information and label semantic information with a specific-designed multi-level feature fusion scheme. During the multi-level feature fusion, an StD mechanism is designed to alleviate the feature collapse caused by network depth.

The structure of the rest of this paper is described as follows: Section 2 introduces the related work from four aspects. Section 3 introduces our proposed hybrid intention recognition framework for intention recognition in the financial domain. Section 4 presents the experiments and discussions, involving comparison experiments, ablation experiments, and error analysis. Finally, Section 5 summarizes this paper, involving future work.

2. Related Work

2.1. Data Augmentation for NLP

In real applications, acquired text data are usually insufficient for deep learning. Data argumentation is an alternative approach to generate sufficient data to train a reliable deep model [20].

Wei et al. [4] provided four easy data augmentation (EDA) techniques for text classification tasks, involving synonym replacement, random insertion, random swap, and random deletion. Experimental results on five public datasets indicated that their EDA method could boost the text classification performance, especially for small datasets. However, the EDA possibly disrupts the semantic consistency in specialized domains (e.g., financial customer service), leading to inappropriate synonym replacements and even grammar errors to in turn degrade the model performance. Sennrich et al. [5] utilized the back-translation method to argument monolingual corpus data for machine translation models. Experimental results on two public datasets indicated that the small amounts of in-domain monolingual data through back-translation could be effectively used for domain-adaptive machine translation. However, they referred that their augmentation method still depended on the amount (and similarity to the test set) of available monolingual data. Moreover, in the financial customer service, only using back-translation easily cause colloquial expressions to lose their original meanings, and back-translating short texts easily results in ambiguity and even erroneous semantics. Schick et al. [6] proposed a Pattern-Exploiting Training (PET) to convert training data into a cloze (fill-in-the-blank) format. Their method fine-tunes a pre-trained language model for each class of samples by using cloze-style phrases and then integrates multiple fine-tuned models to assign soft labels to unlabeled data. Since it leverages the knowledge contained in pre-trained models to generate more accurate labels, their method can be applied to few-shot text classification and natural language inference tasks. Experimental results on five public datasets indicated that the PET could help leverage the knowledge contained within pre-trained language

models for downstream tasks, when the initial amount of training data was limited. However, cloze-style prompts are not well suitable for complex and specialized expressions in financial customer service. Huiming et al. [7] proposed an order-agnostic data augmentation (OADA) method for few-shot named entity recognition (NER). Experimental results on three public NER datasets showed that their method alleviated the one-to-many problem in NER tasks. As they claimed, they augmented the entity data under the assumption that different entity arrangements were equivalent. This assumption is not suitable for intention recognition.

Related data augmentation methods for NLP are summarized in Table 1. Although these data augmentation methods have been validated on some public datasets, the corpus data in financial customer service are quite different from these public datasets. This possibly makes these methods not generate high-quality corpus data, which will degrade the model ability of intention recognition. Especially, augmenting corpus data without corresponding labels possibly limits the generalization ability of the model. Thus, in our study, we will utilize a semi-supervised learning data augmentation method to generate label-flipped corpus data for the improvement of the model's generalization ability [21].

Table 1. A summary of some studies on data augmentation for NLP.

Years	Authors	Methods	Advantages	Limitations
2015	Sennrich et al. [5]	Back-translation.	Increases data diversity.	Easily loses the original meanings of corpus data, and results in ambiguity and even erroneous semantics.
2019	Wei et al. [4]	Synonym replacement, random insertion, random swapping, and random deletion.	Simple and easy implementation.	Possibly disrupts the semantic consistency in specialized domains.
2020	Schick et al. [6]	Pattern-Exploiting Training.	Leverages pre-trained language models.	Cloze-style prompts are not well suitable for complex and specialized expressions.
2024	Huiming et al. [7]	Order-agnostic data augmentation.	Solves the one-to-many problem in NER tasks.	Their assumption is not suitable for intention recognition.

2.2. Convolutional Neural Network Based Methods for NLP

Intention recognition for financial customer service is essentially a NLP task, in which customers' dialogue texts are identified as the pre-defined intention categories [22]. Machine learning for NLP conventionally utilizes the word vectors of the texts constructed by the co-occurrence matrix and global vectors for word representation (GloVe) to train a machine learning model, in which AdaBoost, support vector machine (SVM), and multi-layer perceptron (MLP) are commonly utilized as classifiers [23].

Due to its strong ability of feature mapping via convolutional learning, convolutional neural networks (CNNs) have been widely employed for NLP [24]. Liu et al. [25] integrated three RNNs into a multi-learning framework to conduct joint learning across multiple related text classification tasks about movie reviews, mapping arbitrary text into semantic vector representations with task-specific layers and shared layers. However, when the corpus data are not sufficient enough, the differences between different tasks bring a negative effect for joint learning. Lai et al. [26] used recurrent CNNs to classify text at the document level, including a recurrent structure to capture context information and a max pooling layer to automatically capture key components of text. Although their

method can, to some extent, capture context information, it is weak in capturing the long-term dependencies in the context due to the inherent characteristics of convolutions. Johnson et al. [27] proposed a depth pyramid convolution neural network (DPCNN) for sentiment classification and some other topic classification tasks, in which all the shortcuts were simple identity mapping. However, their proposed text region embedding can only cover one or several words, which cannot well capture the context information in complex expressions.

For these CNNs-based methods, a large amount of corpus data with relatively balanced distributions have been provided to adequately train a promising model for text classification. However, due to confidentiality, it is difficult to acquire massive corpus data in financial customer services scenarios. Also, the quantity of corpus data for some intentions has a high probability of being small.

2.3. Transformer Based Methods for NLP

In recent years, the Transformer architecture [28] has demonstrated strong capabilities in NLP due to its strength of capturing long-term dependencies. Gong et al. [29] proposed a hierarchical graph transformer-based deep learning framework for large-scale multi-label text classification, which cascaded a graph-based model with a transformer encoder architecture. Experiments on three benchmark datasets show that their method can effectively capture the hierarchical structure and logic of the text, as well as the hierarchical relationships of the labels. They claimed that their approach had certain limitations in capturing complex semantics. Lee-Thorp et al. [30] employed basic linear transformations to intermix input tokens and replaced the self-attention mechanism with a Fourier transform to optimize the transformer encoder, which could achieve a par accuracy with BERT at a faster training speed. Due to its Fourier transform, it possibly costs a large amount of time for inferencing and cannot better capture the long-term dependencies compared to the self-attention mechanism.

Büyük et al. [31] utilized transfer learning to fine-tune a pre-trained transformer model for intention recognition, which preliminarily demonstrated the powerful capabilities of the BERT model [14] in the field of intention recognition. Liu et al. [32] proposed the RoBERTa model for text classification, which only trained the BERT model longer with bigger batches to improve the BERT's performance. This simple training strategy does not essentially improve the classification ability of the model. Jiang et al. [33] cascaded a BERT-BiLSTM-TextCNN network for sentiment analysis of netizens' comments, which only transferred the semantics information from shallow to deep without any interactions between shallow and deep. The method by Cai et al. [34] designed a hybrid BERT model for multi-label text classification, which used an attention mechanism to establish the semantic relationship between labels and words. As they claimed, their method would evidently degrade in the case of the imbalanced data with many classes. Similar to [34], Xiong et al. [35] encoded the text of category labels along with documents into the BERT model, which could confuse the correspondence between the label and the text. Benayas et al. [36] fine-tuned the popular pre-trained BERT variant to extract text semantic features for the subsequent joint recognition of intentions and named entities. Although they interacted with the text semantic features for the two tasks, this task-specific interaction possibly influences each other. Shi et al. [37] proposed a protocol regularization training mode for a pre-trained BERT model on high-resource languages to detect offensive language in low-resource languages. Since offensive languages exhibit more distinct features compared to daily languages with colloquial expressions, their method may face some biases when dealing with the colloquial texts. Xu et al. [38] incorporated the auxiliary knowledge extracted by a pre-trained large-

scale language model into the pre-trained BERT for quadruple identification and extraction tasks in the sentiment analysis domain for the tourism sector.

Although these transformer-based methods have performed well on some popular text datasets by fine-tuning pre-trained language models [39,40], the pre-trained models established on general-purpose corpus data tend to have some bias in the characterization learning task in real application scenarios, such as in financial customer service.

Table 2 summarizes some studies on CNN-based and Transformer-based NLP methods. As indicated in Table 2, due to confidentiality issues, it is difficult to acquire a large amount of corpus data in financial customer service scenarios. Additionally, the corpus data for certain intentions may be very scarce, and CNN-based methods [25–27] cannot be adequately trained, which leads to unsatisfactory performance. In contrast, Transformer-based methods can better handle the issue of limited corpus data by fine-tuning pre-trained language models. However, pre-trained models established on general-purpose corpus data are insufficient to address the domain-specific semantics and contexts, which limits the performance improvements of the models.

Table 2. Some studies on CNN-based and Transformer-based methods for NLP.

Years	Authors	Methods	Advantages	Limitations
2015	Lai et al. [26]	TextCNN	The combination of the recursive structure and CNN can enhance the ability to understand the text.	Weak in capturing the long-term dependencies in the context.
2016	Liu et al. [25]	RNN	Shares information between different tasks.	Sufficient corpus data are needed for training.
2017	Johnson et al. [27]	DPCNN	Low computational complexity, and efficiently represents long-term associations in the text.	Cannot well capture the context information in complex expressions.
2019	Liu et al. [32]	RoBERTa	Enhances the model's expressiveness.	Simple training strategy does not essentially improve the classification ability.
2020	Cai et al. [34]	Hybrid BERT	Establishes the semantic relationship between labels and words.	Evidently degrades in the case of the imbalanced data with many classes.
2021	Xiong et al. [35]	Combines label with BERT	Establish the semantic relationship between labels and words.	Confuses the correspondence between the label and the text.
2021	Lee-Thorp et al. [30]	FNet	Replace the self-attention mechanism with Fourier transform to improve training efficiency.	Cannot better capture the long-term dependencies.
2021	Benayas et al. [36]	Fine-tunes the popular pre-trained BERT variant	Enhance the interaction between intention and entity.	Task-specific interactions possibly influence each other.
2022	Shi et al. [37]	Combines protocol regularization training mode with BERT	Strong cross-lingual adaptability.	Weak generalization for colloquial texts.

2.4. Current Development of NLP Methods in the Finance Domain

The development of NLP technology has made it possible to more accurately capture the sentiment and semantics in corpus data, which has been employed in the financial domain. NLP studies in the finance domain can mainly be categorized into traditional corpus-based [41,42] and deep learning-based [43,44] studies, as summarized in Table 3.

Table 3. Some NLP studies in the finance domain.

Years	Authors	Methods	Advantages	Limitations
2019	Araci et al. [43]	Pre-trained BERT	Works well in extracting explicit sentiments.	Does not work well with model-implicit information.
2020	Garcia-Mendez et al. [41]	Corpus-based	The Jaccard distance-based short-text similarity detector reduces training redundancy.	The constructed lexicon and corpus features are designed based on specific rules, and may not be able to cope with complex and rapidly changing environments.
2022	Huang et al. [44]	Pre-trained BERT	Works exceptionally well in identifying positive or negative sentiment.	Does not work well with model-implicit information.
2023	Garcia et al. [42]	Corpus-based	Highlights the more influential words.	The constructed lexicon and corpus features are designed based on specific rules, and may not be able to cope with complex and rapidly changing environments.

Garcia et al. [42] proposed a robust multinomial inverse regression model to construct a lexicon for earning calls between the firm's management and analysts/investors, which could characterize the stock market reaction to the event. Garcia-Mendez et al. [41] proposed a corpus machine learning method combining meta-information and linguistic knowledge for the classifying of short texts in bank transaction descriptions. For text classification via the SVM, the corpus data were preprocessed for linguistic knowledge extraction and selection, successively involving data retrieval, text tokenization and stopwords removal, proper name detection, and sample reduction with Jaccard similarity. The above corpus-based methods highly rely on the quality of the constructed lexicon and corpus features via elaborately selected/designed approaches, which are inflexible in complex and mutable financial environments.

Araci et al. [43] fine-tuned the pre-trained BERT model for financial sentiment analysis. As they claimed, their method could not well model implicit information. Compared to [43], Huang et al. [44] utilized much more financial texts to fine-tune the pre-trained BERT model for sentiment classification, involving three types of financial texts from two websites and a firm's investment database. Since they also directly utilized the pre-trained BERT, their method has a similar problem to the method in [43].

3. Methodologies

In this section, we will provide a detailed introduction of our proposed hybrid framework for intention recognition. Specifically, Section 3.1 briefly provides the architecture of the hybrid framework. Section 3.2 introduces semi-supervised learning data augmentation. The design of label semantics inference is presented in Section 3.3. The design of the text classification network is introduced in detail in Section 3.4.

3.1. Architecture of the Proposed Hybrid Framework

As illustrated in Figure 1, the proposed hybrid framework for intention recognition involves semi-supervised learning data augmentation, semantic inference, and text classification. At the stage of data augmentation, the corpus samples in the training set are augmented with the semi-supervised learning fashion by back-translation and a BERT model. At the stage of semantic inference, label semantics of the corpus samples in the training set are extracted by means of a K-means method. At the stage of text classification, label semantics and the augmented corpus data are incorporated into a two-branch-structure network based on a modified BERT to recognize the financial customers' intentions. The proposed hybrid framework that is well trained by the above stages is evaluated by the test set.

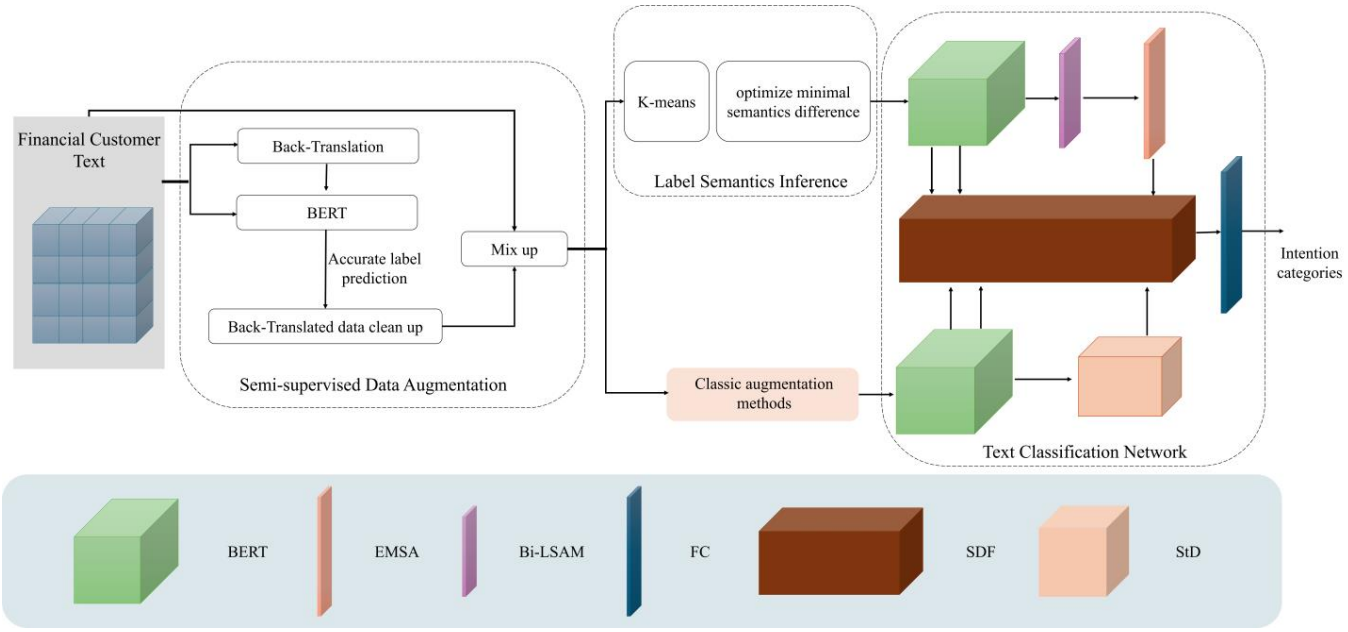

Figure 1. Architecture of the proposed hybrid framework for intention recognition.

3.2. Semi-Supervised Learning Data Augmentation

Due to confidentiality in financial customer service, the acquired corpus data are often insufficient to train an effective text classification network for intention recognition. Semi-supervised learning data augmentation is an alternative to address this issue [45], in which pseudo-labels for unlabeled data generated by a well-trained classifier are utilized to mix pseudo-labeled data with the original data to enrich the original dataset. Inspired by this, a semi-supervised learning method is utilized in this paper to augment the corpus data, involving back-translation and a pre-trained BERT. As illustrated in Figure 1, back-translation is utilized to generate a large amount of low-quality corpus data from the original corpus data. Also, the original corpus data are utilized to fine-tune a pre-trained BERT model. Then, the fine-tuned BERT model is used to clean up the back-translated corpus data. The cleaned corpus data are combined with the original corpus to construct a high-quality corpus dataset.

In this paper, the corpus data of financial customer service are acquired from a financial enterprise in China. Since Japanese and Korean have similar syntaxes with Chinese and even involve a number of Chinese characters, they are utilized as the intermediaries for back-translation. Also, English is used as the intermediary for back-translation, since a number of English characters and digits exist in the corpus data.

Assume that the training set D for the acquired corpus data can be defined as the following:

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{1}$$

where $x_i (i = 1, 2, \ldots, n)$, and y_i denote the i-th corpus sample and the corresponding ground-truth category label, respectively. n denotes the number of corpus samples for training. Then, the low-quality augmented corpus dataset can be acquired by back-translation, defined as follows:

$$\begin{aligned} D_a = \{ &(x'_{11}, y_1), (x'_{12}, y_1), \ldots, (x'_{1m}, y_1), (x'_{21}, y_2), (x'_{22}, y_2), \ldots, \\ &(x'_{2m}, y_2), \ldots, (x'_{n1}, y_n), (x'_{n2}, y_n), \ldots, (x'_{nm}, y_n)\} \end{aligned} \tag{2}$$

where $x'_{ij} (j = 1, 2, \ldots, m)$ denotes the new corpus data generated by x_i. m denotes the number of augmented samples for each x_i.

Since some newly generated corpus data cannot fully reflect the original semantics of the corpus sample due to the influence of linguistic paraphrase diversity [46], the low-quality augmented corpus data possibly influence the generalization ability of the text classification model trained by these data. To obtain high-quality corpus data, a pre-trained BERT model [14] is fine-tuned by the original acquired corpus data to clean up the low-quality augmented corpus data. To do so, the pre-trained BERT model is fine-tuned on the training set D to predict category labels for the back-translated corpus dataset D_a, achieving a new dataset D_a^* with predicted category labels that are obtained as follows:

$$\begin{aligned} D_a^* = \{ &(x'_{11}, y'_{11}), (x'_{12}, y'_{12}), \ldots, (x'_{1m}, y'_{1m}), (x'_{21}, y'_{21}), (x'_{22}, y'_{22}), \ldots, \\ &(x'_{2m}, y'_{2m}), \ldots, (x'_{n1}, y'_{n1}), (x'_{n2}, y'_{n2}), \ldots, (x'_{nm}, y'_{nm})\} \end{aligned} \tag{3}$$

where y'_{ij} denotes the predicted label of the back-translated corpus sample x'_{ij}, predicted by the fine-tuned BERT model. Then, the corpus data that their predicted category labels y'_{ij} that are consistent with the ground-truth ones y_i are manually guaranteed to involve corresponding correct semantics. These guaranteed corpus data are combined with the original acquired corpus data to construct the high-quality corpus dataset.

3.3. Label Semantic Inference

In financial service scenarios, many corpus texts usually consist of 3 to 13 words to express the customers' intentions, which possibly involve highly similar expressions with different semantics (i.e., different semantic labels). Such short corpus texts will influence text classification, if they are directly utilized for classification. Thus, we aim to integrate semantic label-related knowledge into the feature information of the corpus texts. Here, we assume that it is sufficient for this integration of semantic label-related knowledge to identify an approximately central text of the label semantics. Due to its advantages of having a small number of parameters, a fast convergence, and effective clustering, K-means clustering is utilized for this identification of the approximately central text for each semantic category in the short-text corpus data. Our purpose is to find a representative corpus text in a category of corpus texts with similar semantics to characterize this category, rather than to simultaneously cluster all the corpus texts. Thus, for simplicity, the value of k for K-means is set to 1 for each semantic label category. Figure 2 illustrates the process of label semantic inference.

Firstly, the corpus samples are divided into words using jieba (0.42.1). Then, the vector space model (VSM) [47] is constructed according to term frequency–inverse document frequency (TF-IDF) [48] defined as the following:

$$W = TF * IDF \tag{4}$$

where *TF* and *IDF* denote term frequency and inverse document frequency, respectively. $*$ denotes the product of the element. To achieve the center of the category, the K-means is utilized to cluster the corpus samples. To find the approximately central text that represents the label semantics, K-means is utilized to cluster the corpus data, formulated as follows:

$$V = \sum_{v=0}^{A-1} \sum_{S_u \in C_v} (S_u - Z)^2 \tag{5}$$

where A denotes the number of classes in the corpus dataset (here A = 49). S_u denotes all the corpus samples in the category, and C_v, and Z denotes the category center of C_v, V denotes the sum of distances between all corpus samples S_u in the C_v category and the category center of C_v. Semantic inference for the label of each corpus sample can be transformed into the problem of optimizing the minimal semantic difference L between the corpus sample S and the category center Z, defined as follows:

$$L = min(dist(\boldsymbol{S}, Z)) \tag{6}$$

where $dist(\boldsymbol{S}, Z)$ denotes the Euclidean distance of the corpus sample $\boldsymbol{S}$ and the label center Z. The corpus sample with the minimal semantic difference L is set as the label semantics of the category to which it belongs.

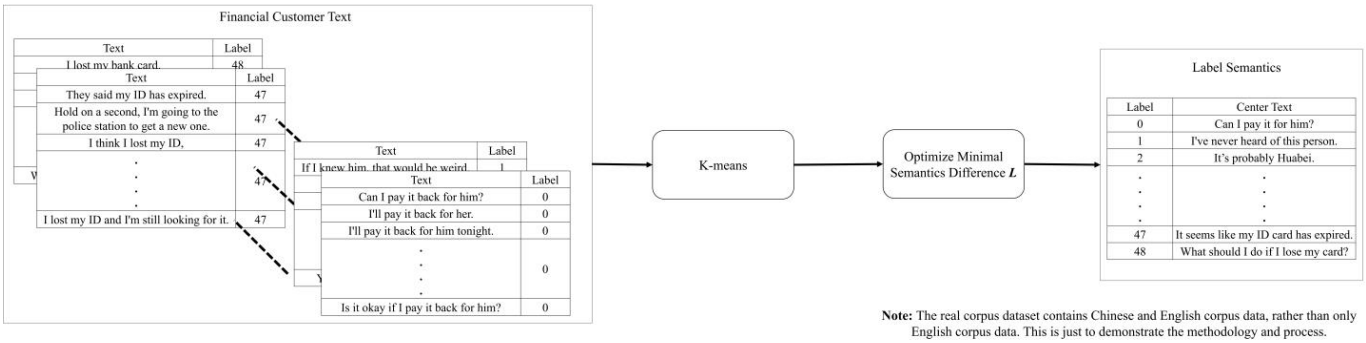

Figure 2. A K-means-based scheme for semantics inference.

3.4. Text Classification Network

As mentioned in Section 3.1, the integration of the label semantics is beneficial for text classification. So, we designed a dual-branch network architecture for text classification, in which label semantics and corpus data are processed by a label semantics information branch and a corpus information branch, respectively, and are integrated by the designed shallow and deep Fusion (SDF) module. As illustrated in Figure 3, the designed text classification network is composed of two BERT encoders for shallow feature extraction, a cascading combination of the BiLSTM [49] and the efficient multi-head self-attention (EMSA) [50] for deep feature extraction, an an SDF module for shallow and deep feature fusion and an StD mechanism for the alleviation of feature collapse.

Specifically, since the bidirectional Transformer encoder of the BERT model can effectively capture contextual dependencies and has a good generalization ability for NLP tasks, two pre-trained BERT models [14], each of which is composed of 12 BERT layers, is utilized in the shallow layers of the network to separately extract the features from the label semantics and the corpus data. With the feature transferring layer by layer, feature collapse easily occurs in the Transformer architecture, which can be addressed by residual connections [51]. Moreover, we attempted to integrate the label semantics as label prior knowledge into the corpus data to improve the classification ability. So, an StD mechanism is proposed to alleviate the feature collapse in the corpus information branch rather than

in the label semantics information branch. To effectively extract the sequence information embedded in the label semantics and to reduce network parameters, the Bi-LSTM and the EMSA are cascaded in the deep layer of the label semantics information branch to well extract deep label semantic features. In the BERT model, lower layers mainly learn the basic features of the text, such as word spelling patterns and basic grammatical structures; middle layers can capture further syntactic relations and phrase structures; while deeper layers mainly extract deep semantic information and complex contextual dependencies [52]. Inspired by this idea, the SDF module is designed to merge the label semantics and corpus data at the levels of word spelling patterns and basic grammatical features, syntactic features, and semantic features. Moreover, in the SDF module, these fused features at the three levels are concatenated to help the classification network well understand the semantics of short texts.

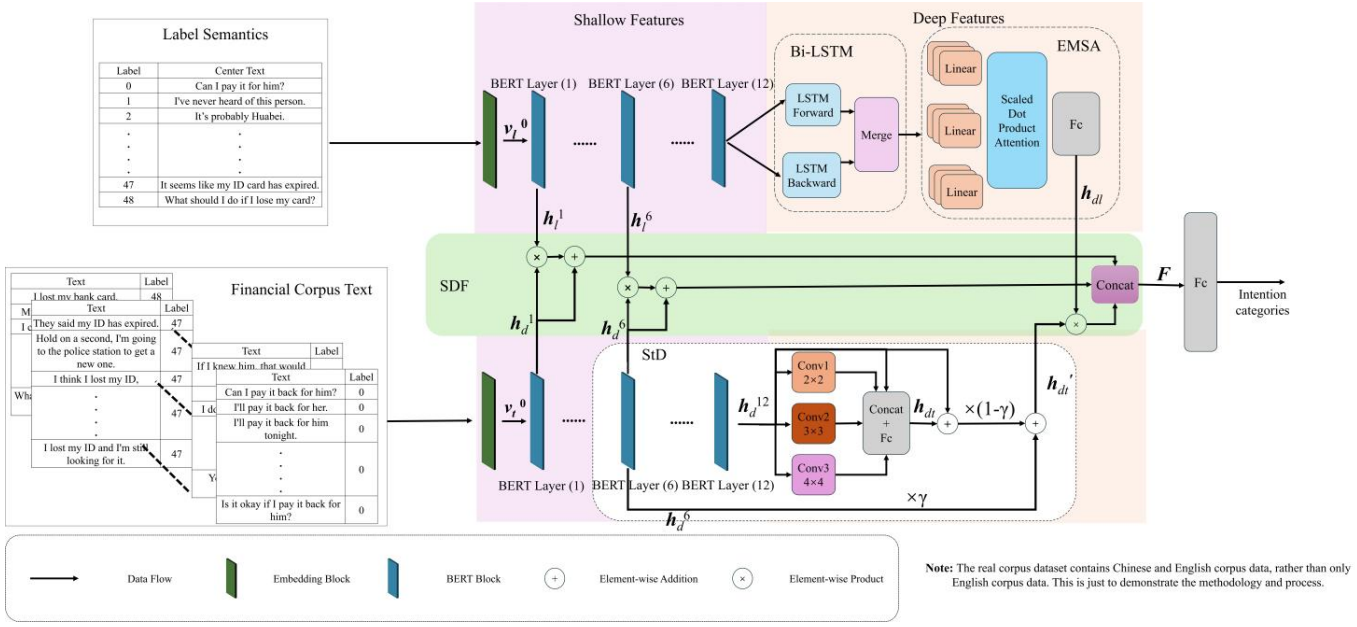

Figure 3. Text classification network.

Assume the input of the i-th BERT layer in the corpus information branch is $\boldsymbol{v}_t^{i-1}$, its corresponding output $\boldsymbol{h}_t^i$ can be formulated as follows:

$$\boldsymbol{h}_t^i = BERT^i\left(\boldsymbol{v}_t^{i-1}\right) \tag{7}$$

where $BERT^i(\cdot)$ denotes the operation of the i-th BERT layer. It is noted that $\boldsymbol{v}_t^0$ is the original corpus data. Then, the StD mechanism is proposed to perform the multi-scale fusion of shallow corpus information and its corresponding deep information, involving multiple skip connections, three convolutions with different kernels, and a combination operation of a concatenation and fully connection (FC), as illustrated in Figure 4. Thus, the deep features of the corpus data can be formulated as follows:

$$\boldsymbol{h}'_{dt} = (1-\gamma) \times (\boldsymbol{h}_{dt} + \boldsymbol{h}_t) + \gamma \times \boldsymbol{h}_t^6 \tag{8}$$

$$\boldsymbol{h}_{dt} = FC\left(concat\left(\boldsymbol{h}_t^{12}, conv1\left(\boldsymbol{h}_t^{12}\right), conv2\left(\boldsymbol{h}_t^{12}\right), conv3\left(\boldsymbol{h}_t^{12}\right)\right)\right) \tag{9}$$

where $\gamma \in (0, 1]$ is a trainable parameter, which can help alleviate feature collapse and increase feature diversity. $concat(\cdot)$ is a concatenating operation. The kernel sizes of the three convolutions, i.e., $conv1(\cdot)$, $conv2(\cdot)$, and $conv3(\cdot)$, are set to (2, 768), (3, 768), and (4, 768), correspondingly.

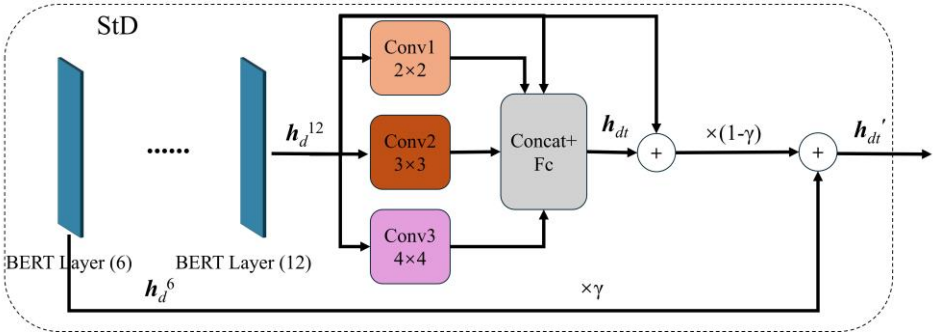

Figure 4. StD mechanism.

Assume the input of the i-th BERT layer in the label semantic information branch is $\boldsymbol{v}_l^{i-1}$, its corresponding output $\boldsymbol{h}_l^i$ can be formulated as follows:

$$\boldsymbol{h}_l^i = BERT^i\left(\boldsymbol{v}_l^{i-1}\right) \tag{10}$$

It is noted that $\boldsymbol{v}_l^0$ is the inferenced label semantics. Next, the encoded label semantic features $\boldsymbol{h}_l^{12}$ via the BERT encoder successively pass through the LSTM and the EMSA, formulated as the following:

$$\boldsymbol{h}_{dl} = EMSA\left(BiLSTM\left(\boldsymbol{h}_l^{12}\right)\right) \tag{11}$$

In (11), BiLSTM(·) denotes the operation of the BiLSTM whose hidden layer dimension is set to 1024. EMSA(·) denotes the operation of the EMSA, in which the number of the multi-head attention heads is set to 4 and the hidden layer dimension is set to 852.

To better characterize the corpus data, the extracted shallow and deep features are fused as $\boldsymbol{H}$, and formulated as follows:

$$\boldsymbol{H} = concat\left(\left(\boldsymbol{h}_t^1 * \boldsymbol{h}_l^1 + \boldsymbol{h}_t^1\right), \left(\boldsymbol{h}_t^6 * \boldsymbol{h}_t^6 + \boldsymbol{h}_t^6\right), \left(\boldsymbol{h}'_{dt} * \boldsymbol{h}_{dl}\right)\right) \tag{12}$$

where $\left(\boldsymbol{h}_t^1 * \boldsymbol{h}_l^1\right)$ and $\left(\boldsymbol{h}_t^6 * \boldsymbol{h}_t^6\right)$ are the integrated shallow features of the corpus data and label semantics at the first and sixth BERT layers, respectively. $\left(\boldsymbol{h}'_{dt} * \boldsymbol{h}_{dl}\right)$ denotes the corresponding integrated deep features. Finally, the features $\boldsymbol{H}$ pass through the FC layer to implement text classification.

4. Experimental Results and Discussions

4.1. Dataset and Training Details

The corpus data of financial customer service were provided by a financial company in China, involving 2977 corpus texts about loan service. Figure 5 illustrates the data distribution of the original corpus data, with the label indices range from 0 to 48. There is a certain discrepancy in the data distribution across various categories, with the ratio of the largest category size to the smallest one being approximately 3:1. Moreover, as shown in Figure 6, the lengths of the original corpus data primarily range from 3 to 20 words. Also, the corpus samples contain a mixture of Chinese and English, which can easily lead to semantic confusions. To well train the text classification network, the semi-supervised data augmentation method described in Section 3.2 is utilized to augment the original corpus texts to 3904 ones, which is split into training/validation/test sets in a ratio of 2688:384:832. Then, 2688 corpus texts are further preprocessed by several traditional augmentation methods to construct the final training set in this study, involving random deletion (twice), synonym replacement (twice), and back-translation (once). That is, 16,128 augmented

corpus texts via semi-supervised data augmentation and traditional data augmentation were utilized for training, while 384 and 832 ones via semi-supervised data augmentation for validation and testing were utilized.

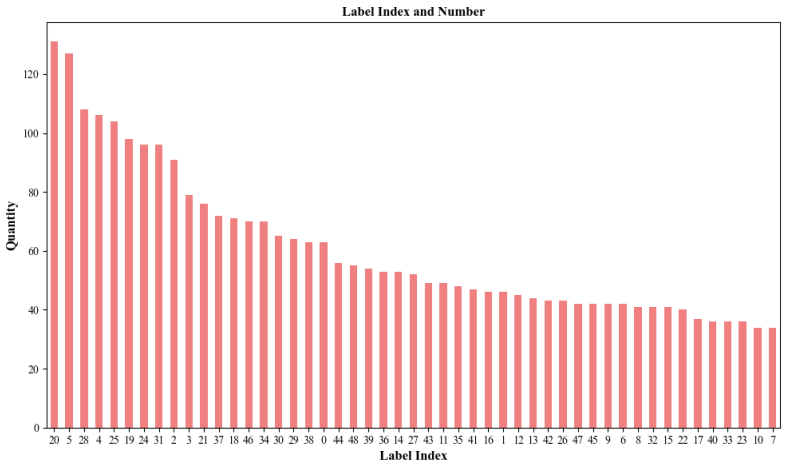

Figure 5. Label indexes and the corresponding counts for an original set.

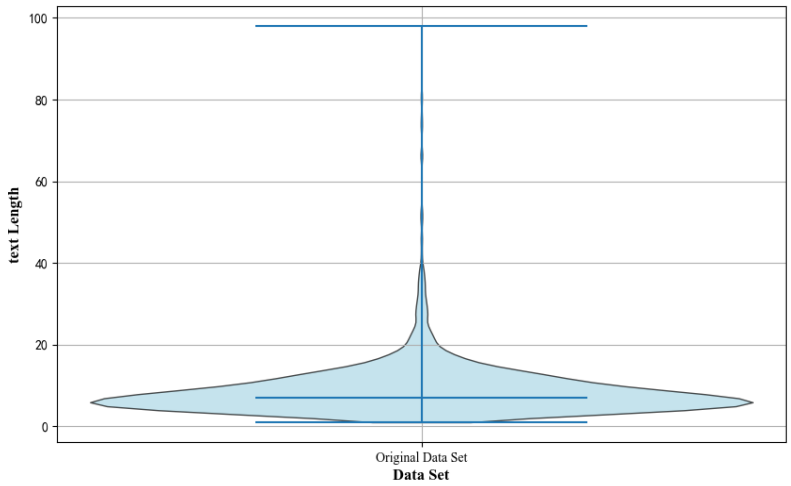

Figure 6. A violin plot of corpus sample lengths.

All the experiments were conducted on a computer with an intel(R)Xeon(R)CPU Es-2630 v3 @ 2.40GHz, a 64G RAM, NVIDIA GeForce RTX 2080 Ti. Python3 was utilized as the programming language, and PyTorch as the deep learning framework due to its widespread use in the field of deep learning. The pre-training model used in the experiments was the bert-base-chinese model with the following parameters: the batch_size was set to 8, the max_len was set to 30 for samples, the max_len was set to 400 for labels, the loss function was the cross-entropy loss function, the optimizer was Adam, the learning rate was set to 2×10^{-5}, every 10 epochs was multiplied by a 0.1 attenuation factor, and the epoch was set to 30. Four commonly used metrics were utilized to perform a comprehensive evaluation of the intention recognition, which accuracy (Acc), precision (Prec), Recall, and F1-score (F1).

4.2. Comparisons with Other Deep Learning Methods

To validate the proposed hybrid intention recognition framework, we utilized eight deep learning-based text classification methods for comparisons, involving two CNN-based models (TextRCNN [26] and DPCNN [27]) and six Transformer-based models (FNet [30], BERT [14], Xiong et al. [35], FinBERT [44], BERT-DBRU [53], and CBLMA-B [54]). Specifically, there are five models specifically designed for text classification, involving two

CNN-based models, a Transformer-structure model (FNet), and two BERT-based models (BERT [14] and Xiong et al. [35]). Two BERT-based models are specifically designed for intention recognition, involving BERT-DBRU and CBLMA-B. FinBERT is a tailored fine-tune BERT model for the financial domain, demonstrating strong capabilities in sentiment classification for financial texts. During training, the two CNN-based models were directly trained on our training set. FNet and FinBERT were fine-tuned on our training set using their pre-trained models. The other four BERT-based models were fine-tuned on our training set, using the pre-trained Google BERT model [14]. For fair comparisons, all the comparison models were trained according to the configurations described in their studies and tested on our test set. The comparison results achieved by different methods are shown in Table 4.

Table 4. Comparison results achieved by different deep learning-based methods.

Methods	Acc (%)	Prec (%)	Recall (%)	F1 (%)	Inference Time/s
TextRCNN (2015) [26]	79.59	80.33	81.15	80.02	0.62
DPCNN (2017) [27]	68.30	70.08	68.94	67.91	**0.36**
BERT (2019) [14]	85.52	87.89	86.15	86.30	0.86
Xiong et al. (2021) [35]	85.75	87.67	87.40	86.94	1.48
FNet (2021) [30]	87.50	88.34	88.49	88.08	12.65
FinBERT (2022) [44]	86.78	87.78	88.29	87.68	0.80
BERT-DBRU (2023) [53]	85.70	87.70	86.89	86.67	1.64
CBLMA-B (2024) [54]	87.50	88.68	88.63	88.19	4.65
Ours	**89.06**	**90.27**	**90.40**	**90.07**	11.87

The DPCNN [27] increases the perceptual field by multiple down-sampling operations to obtain the long-term dependency of the text. However, the associations of phrases within the text need to be supported by sufficient data, and customers with the same intention category have different expressions in real scenarios. These two facts result in the fact that the DPCNN cannot function well for new text expressions.

The TextRCNN [26] utilizes the RNN to capture the contextual information of the text, which results in a much better intention recognition performance than the DPCNN. However, the TextRCNN is also a one-hot encoder method, which limits its ability to extract the associations of word-to-word. Moreover, the short length of the sentences in financial service scenarios leads to limited contextual information in the text.

FNet [30] replaces the self-attention layer in traditional Transformer networks with Fourier transform to capture feature information by transforming the features into the frequency domain. So, it can achieve a much better intention recognition performance than the two CNN-based models and several BERT models. However, its Fourier transform possibly loses some long-term dependencies, which to some extent influences its recognition ability.

Due to BERT's ability to effectively capture global contextual semantic representations, five existing BERT-based methods (BERT [14], Xiong et al. [35], FinBERT [44], BERT-DBRU [53], and CBLMA-B [54]) perform significantly better than the two CNN-based models. Xiong et al. [35] cascades the BERT structure and the bi-GRUs to promote the pre-trained Google BERT model, while BERT-DBRU [53] concatenates texts and corresponding text labels to input into the pre-trained Google BERT model. So, the two models more or less perform a better intention recognition than the Google BERT [14]. Since FinBERT [44] directly fine-tunes the Google BERT model on a large amount of financial text data, it can achieve a fairly good performance for intention recognition on our dataset, with an accuracy of 86.78%, a precision of 87.78%, a recall of 88.29%, and an F1 score of 87.68%. Comparatively, CBLMA-B [54] utilizes a relatively complex network structure for feature extraction,

which is a dual-branch structure incorporating BERT, CNN, BiLSTM, and self-attention. So, it achieves the second-best performance for intention recognition, with an accuracy of 87.50%, a precision of 88.68%, a recall of 88.63%, and an F1 score of 88.19%. However, its simple concatenation of the dual-branch features cannot reveal the inherent relationship between texts and corresponding text labels, which limits its intention recognition ability.

Our model performs best in the intention recognition task for financial customer service, achieving an accuracy of 89.06%, a precision of 90.27%, a recall of 90.40%, and an F1 of 90.07%. This is primarily attributed to four aspects. Firstly, the label semantic inference scheme provides effective label information for datasets lacking real label semantics and serves as a constraint for subsequent text classification. Secondly, the text classification network seamlessly integrates shallow and deep features of the corpus data and their label semantics at different levels for intention recognition. Thirdly, the StD mechanism enriches the semantic features in the deep network, alleviating the issue of feature collapse. Additionally, the augmented dataset provides more trainable data, enabling the model to learn more variations in the intention categories.

Tabel 4 illustrates the inference time for 832 corpus texts via various models. Two CNN-based models have significant advantages in terms of inference speed due to their simple structures and small numbers of parameters. The Fourier transforms in FNet are time-consuming, which results in the most inference time of 12.65 s for FNet. Since the three BERT-based models (i.e., Xiong et al. [35], FinBERT [44], and BERT-DBRU [53]) utilize various simple schemes to improve the classification ability of the Google BERT model [14], they cost nearly inference time, ranging from 0.8 s to 1.7 s. In comparison, the dual-branch structure of CBLMA-B [54] makes it time-consuming, with an inference time of 4.65 s. Due to feature interactions in the dual-branch structure, our model consumes more inference time than CBLMA-B [54], with an inference time of 11.87 s.

4.3. Ablation Experiments

Here, we conducted ablation experiments to discuss the influences of the semi-supervised data augmentation (SDA) and label semantic inference (LSI) on the hybrid framework, and those of the network structures on the text classification network. Since SDA and LSI generate the augmented corpus data and corresponding label semantics, respectively, which serve as the inputs of the text classification network, they do not participate in the joint training of the text classification network. So, we conducted two ablation experiments to separately discuss the influences of SDA and LSI, as well as those of the text classification network structures, as shown in Tables 5 and 6, respectively.

Table 5. Framework ablation study.

Baseline	SDA	LSI	Acc (%)	Prec (%)	Recall (%)	F1 (%)
√			85.52	86.24	87.40	86.42
√	√		86.66	87.82	88.00	87.59
√		√	86.20	87.65	87.70	87.21
√	√	√	**89.06**	**90.27**	**90.40**	**90.07**

Table 6. Module ablation study.

Baseline	SDF	StD	Acc (%)	Prec (%)	Recall (%)	F1 (%)
√			87.80	89.09	89.57	88.95
√	√		88.94	89.37	90.08	89.48
√		√	88.46	89.62	89.66	89.39
√	√	√	**89.06**	**90.27**	**90.40**	**90.07**

4.3.1. Framework Ablation Study

To validate SDA and LSI in our hybrid framework, we conducted an ablation experiment to discuss the text classification network (baseline model) with/without SDA and/or LSI. The hybrid framework only with SDA shows that the inputs of the baseline model are the original corpus data and their corresponding label semantics. The hybrid framework only with LSI shows that the inputs of the baseline model are the augmented corpus data.

As shown in Table 5, the hybrid framework with only LSI performs better than the baseline model, indicating that the incorporation of label semantic knowledge into the corpus data can effectively help the text classification network learn label-related semantic information to alleviate the influence of ambiguous expressions. However, insufficient corpus data make LSI not well find the approximately central text representing the label semantics. Thus, the hybrid framework with only LSI is inferior to that with only SDA, with an intention recognition performance of 87.21% vs. 87.59% F1, demonstrating the significance of sufficient corpus data for intention recognition. When both SDA and LSI are combined with the baseline model, the hybrid framework achieves the best intention recognition performance.

4.3.2. Classification Module Ablation Study

To validate SDF and StD in our designed text classification network, we conducted an ablation experiment to discuss the hybrid framework with/without SDF and/or StD. Specifically, the baseline framework in this ablation study shows that the hybrid framework involves SDA and LSI for augmented corpus data and the corresponding label semantics, and the dual-branch baseline text classification model without SDF and StD.

Since StD can, to some extent, alleviate the feature collapse with the increase of the network depth to enrich the representation of deep features, the hybrid framework with only StD performs better than the baseline framework, with an intention recognition performance of 89.39% vs. 88.95% F1. However, due to the lack of effective interactions between label semantics and corpus semantics, the text classification network fails to effectively integrate label semantics into the corpus information. Therefore, the text classification network using only StD is less effective in intention recognition performance compared to the one using only SDF, with F1s of 89.48% and 89.39%, respectively. This indicates the importance of label semantic fusion for intention recognition. When the baseline framework is combined with both SDF and StD, its corresponding intention recognition performance reaches the best value.

4.4. Analysis

To further demonstrate the intention recognition ability of our proposed hybrid framework, we have illustrated the confusion matrix of the predicted results for the test set in Figure 7.

As indicated in Figure 7, most of the corpus texts are correctly recognized in their intentions by our hybrid framework, with an 89.06% accuracy. However, due to complicated financial text expressions, some intentions are misunderstood. For example, some confusions occur between the corpus data with Label 24 and those with Labels 4, 5, 8, and 25. After we checked the related corpus data, we found out that exactly that some corpus data have ambiguous short expressions which result in these confusions. For instance, the text "No problem." with Label 24 is semantically similar to the text "No problem. I will do it." with Label 4; the text "I will repay it." with Label 24 vs. the text "I will repay it today." with Label 5; the text "It's wrong in time." with Label 24 vs. the text "No, it's wrong." with Label 8; and the text "I will repay it." with Label 24 vs. the text "I will repay it next month."

with Label 25. In the future, we will collect more financial corpus data to better learn the subtle differences among various corpus texts.

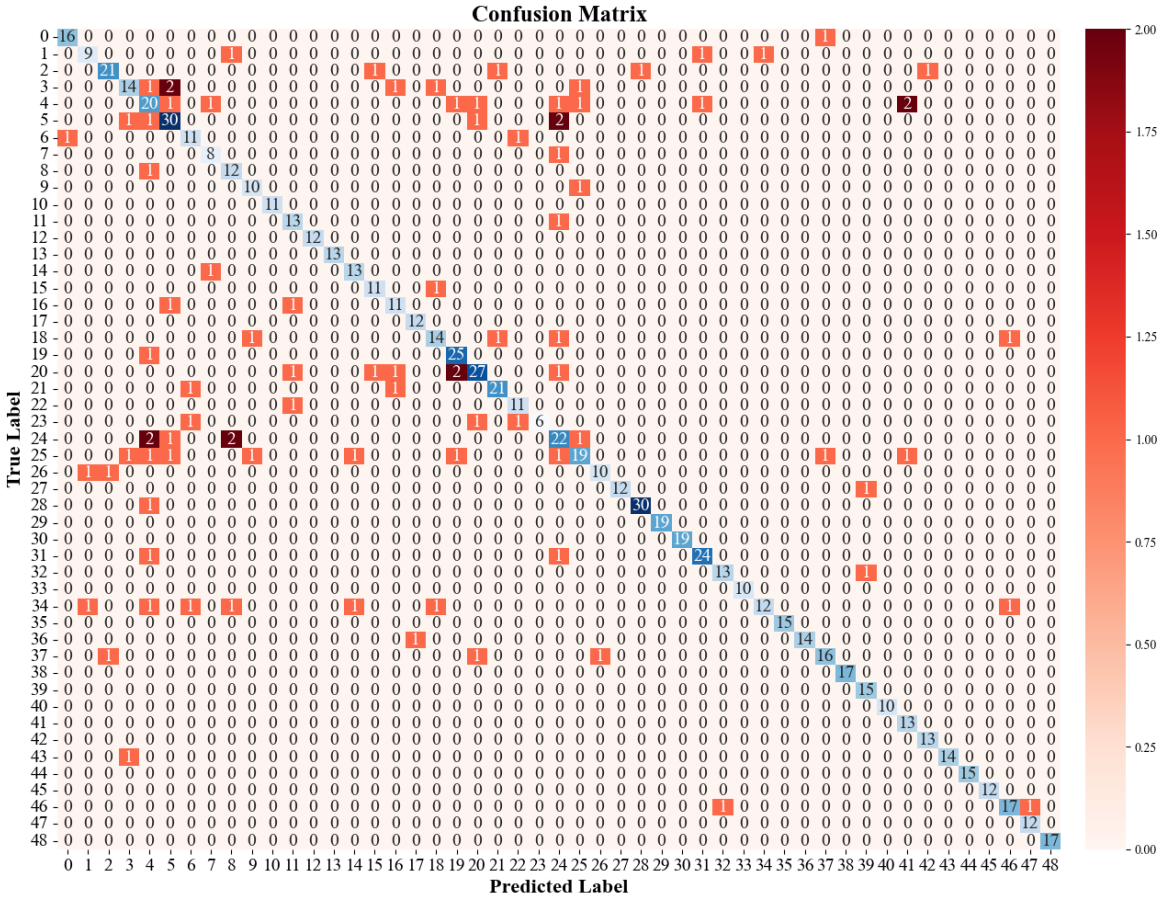

Figure 7. Confusion matrix for the results.

5. Conclusions

Due to the small number of samples, high colloquialism, and semantic ambiguity, automatic intention recognition is a challenging task in finance customer service. In this paper, a hybrid framework is designed for automatically recognizing financial customer intentions, consisting of semi-supervised learning data augmentation for high-quality corpus data acquisition, K-means-based label semantic inference for labels constraints of intention recognition, and a specific-designed BERT-based text classification. Here, a dual-branch BERT deep network is designed, which integrates corpus information and label semantic information, and an StD mechanism for alleviating semantic feature collapse. Compared to existing deep learning methods, the hybrid framework achieves better in financial customer service intention recognition on a corpus dataset about loan service provided by a financial company in China, with an accuracy of 89.06%, a precision of 90.27%, a recall of 90.40%, and an F1 score of 90.07%. The facts demonstrate the potential of the hybrid framework to automatic intention recognition in financial customer service, enabling the automatic identification of customer intentions to further provide timely and accurate services. Moreover, our method can reduce the manual intervention to save labor costs.

Despite the excellent performance of the framework in intention recognition, more than 10% of the samples in the test set were not correctly recognized. This may be due to the limited number of training samples and the framework's inability to effectively extract emotional contextual semantics from these samples. In the future, we can further

fine-tune the model by collecting a large amount of corpus data related to NLP tasks in the financial domain to better learn the subtle differences between various corpus texts, which can enhance the model's ability to recognize colloquial texts in the financial domain, and provide stronger support for intelligent financial customer service.

Author Contributions: Conceptualization, N.C.; methodology, N.C. and S.L.; software, J.X. and S.L.; validation, N.C. and S.L.; formal analysis, N.C.; investigation, S.L.; resources, J.X. and Y.T.; data curation, S.L. and Y.T.; writing—original draft preparation, N.C. and S.L.; writing—review and editing, N.C.; visualization, J.X. and Y.T.; supervision, Y.Z. and J.L.; project administration, Y.Z. and J.L.; funding acquisition, Y.Z. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Science and Technology Plan Project of Maoming City, Guangdong Province (2022DZXHT034, 2022S048).

Institutional Review Board Statement: This retrospective study and the included supervised procedures were approved by PSBC Consumer Finance's consent.

Informed Consent Statement: Not applicable.

Data Availability Statement: It should be noted that the data collected for research purposes is subject to informed PSBC Consumer Finance's consent, but due to confidentiality agreements, the dataset does not support open access at this time. If you would like to obtain some data, please contact cainian@gdut.edu.cn.

Acknowledgments: This work has received support from Guangdong University of Technology and PSBC Consumer Finance. We would like to thank the anonymous reviewers.

Conflicts of Interest: Author Jiacheng Liao was employed by the company Foshan Talkiin Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

Acc	accuracy
BERT	bidirectional encoder representation from transformers
CNNs	convolutional neural networks
DPCNN	depth pyramid convolution neural network
EDA	easy data augmentation
EMSA	efficient multi-head self-attention
FC	fully connection
F1	F1-score
GloVe	global vectors for word representation
IoT	Internet of Things
LSI	label semantic inference
MLP	multi-layer perceptron
NLP	natural language processing
NER	named entity recognition
OADA	order-agnostic data augmentation
PET	Pattern-Exploiting Training
Prec	precision
StD	shallow-to-deep
SVM	support vector machine
SDF	shallow and deep Fusion
SDA	semi-supervised data augmentation
TF-IDF	term frequency–inverse document frequency
VSM	vector space model

References

1. Lin, J.; Yu, W.; Zhang, N.; Yang, X.; Zhang, H.; Zhao, W. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet Things J.* **2017**, *4*, 1125–1142. [CrossRef]
2. Ponce, V.; Abdulrazak, B. Intention as a context: An activity intention model for adaptable development of applications in the internet of things. *IEEE Access* **2021**, *9*, 151167–151185. [CrossRef]
3. Kyj, L.S.; Kyj, M.J. Customer service: Product differentiation in international markets. *Int. J. Phys. Distrib. Mater. Manag.* **1989**, *19*, 30–38. [CrossRef]
4. Wei, J.; Zou, K. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv* **2019**, arXiv:1901.11196.
5. Sennrich, R.; Haddow, B.; Birch, A. Improving neural machine translation models with monolingual data. *arXiv* **2015**, arXiv:1511.06709.
6. Schick, T.; Schütze, H. Exploiting cloze-questions for few-shot text classification and natural language inference. In Proceedings of the European Chapter of the Association for Computational Linguistics, Online, 5–10 July 2020.
7. Wang, H.; Cheng, L.; Zhang, W.; Soh, D.W.; Bing, L. Order-Agnostic Data Augmentation for Few-Shot Named Entity Recognition. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 11–16 August 2024.
8. Cui, L.; Huang, S.; Wei, F.; Tan, C.; Duan, C.; Zhou, M. Superagent: A customer service chatbot for e-commerce websites. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017.
9. Rabino, S.; Onufrey, S.R.; Moskowitz, H. Examining the future of retail banking: Predicting the essentials of advocacy in customer experience. *J. Direct, Data Digit. Mark. Pract.* **2009**, *10*, 307–328. [CrossRef]
10. Zou, Y.; Liu, H.; Gui, T.; Wang, J.; Zhang, Q.; Tang, M.; Li, H.; Wang, D. Divide and conquer: Text semantic matching with disentangled keywords and intents. *arXiv* **2022**, arXiv:2203.02898.
11. Chen, F.; Li, M.; Wu, H.; Xie, L. Web service discovery among large service pools utilising semantic similarity and clustering. *Enterp. Inf. Syst.* **2017**, *11*, 452–469. [CrossRef]
12. Yang, Z.; Wang, L.; Wang, Y. Multi-intent text classification using dual channel convolutional neural network. In Proceedings of the 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 6–8 June 2019; pp. 397–402.
13. Goyal, P.; Pandey, S.; Jain, K. *Deep Learning for Natural Language Processing*; Springer: Cham, Switzerland, 2019.
14. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2–7 June 2019.
15. Devi, S.A.; Ram, M.S.; Dileep, P.; Pappu, S.R.; Rao, T.S.M.; Malyadri, M. Positional-attention based bidirectional deep stacked AutoEncoder for aspect based sentimental analysis. *Big Data Res.* **2025**, *39*, 100505. [CrossRef]
16. Shah, P.; Patel, H.; Swaminarayan, P. Multitask Sentiment Analysis and Topic Classification Using BERT. *ICST Trans. Scalable Inf. Syst.* **2024**, *11*, 1–12. [CrossRef]
17. Cheng, Q.; Shi, W. Hierarchical multi-label text classification of tourism resources using a label-aware dual graph attention network. *Inf. Process. Manag.* **2025**, *62*, 103952. [CrossRef]
18. Anno, S.; Kimura, Y.; Sugita, S. Using transformer-based models and social media posts for heat stroke detection. *Sci. Rep.* **2025**, *15*, 742. [CrossRef] [PubMed]
19. Petrillo, L.; Martinelli, F.; Santone, A.; Mercaldo, F. Explainable Security Requirements Classification Through Transformer Models. *Futur. Internet* **2025**, *17*, 15. [CrossRef]
20. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 63. [CrossRef]
21. Zhou, J.; Zheng, Y.; Tang, J.; Jian, L.; Yang, Z. FlipDA: Effective and Robust Data Augmentation for Few-Shot Learning. *arXiv* **2021**, arXiv:2108.06332.
22. Guo, S.; Wang, Q. Application of knowledge distillation based on transfer learning of ernie model in intelligent dialogue intention recognition. *Sensors* **2022**, *22*, 1270. [CrossRef]
23. Li, Q.; Peng, H.; Li, J.; Xia, C.; Yang, R.; Sun, L.; Yu, P.S.; He, L. A survey on text classification: From traditional to deep learning. *ACM Trans. Intell. Syst. Technol.* **2020**, *13*, 31. [CrossRef]
24. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning-based text classification. *ACM Comput. Surv.* **2020**, *54*, 40. [CrossRef]
25. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classification with multi-task learning. *arXiv* **2016**, arXiv:1605.05101.
26. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; Volume 29. [CrossRef]
27. Johnson, R.; Zhang, T. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the Annual Meeting of the Association for Computational Linguistics 2017, Vancouver, BC, Canada, 30 July–4 August 2017.

28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
29. Gong, J.; Ma, H.; Teng, Z.; Teng, Q.; Zhang, H.; Du, L.; Chen, S.; Alam Bhuiyan, Z.; Li, J.; Liu, M. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification. *IEEE Access* **2020**, *8*, 30885–30896. [CrossRef]
30. Lee-Thorp, J.; Ainslie, J.; Eckstein, I.; Ontanon, S. Fnet: Mixing tokens with fourier transforms. *arXiv* **2021**, arXiv:2105.03824.
31. Buyuk, O.; Erden, M.; Arslan, L.M. Leveraging the information in in-domain datasets for transformer-based intent detection. In Proceedings of the 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), Elazig, Turkey, 6–8 October 2021; pp. 1–4.
32. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. toRoBERTa: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
33. Jiang, X.; Song, C.; Xu, Y.; Li, Y.; Peng, Y. Research on sentiment classification for netizens based on the bert-bilstm-textcnn model. *PeerJ Comput. Sci.* **2022**, *8*, e1005. [CrossRef] [PubMed]
34. Cai, L.; Song, Y.; Liu, T.; Zhang, K. A hybrid bert model that incorporates label semantics via adjustive attention for multi-label text classification. *IEEE Access* **2020**, *8*, 152183–152192. [CrossRef]
35. Xiong, Y.; Feng, Y.; Wu, H.; Kamigaito, H.; Okumura, M. Fusing label embedding into bert: An efficient improvement for text classification. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 1–6 August 2021.
36. Benayas, A.; Hashempour, R.; Rumble, D.; Jameel, S.; de Amorim, R.C. Unified Transformer Multi-Task Learning for Intent Classification with Entity Recognition. *IEEE Access* **2021**, *9*, 147306–147314. [CrossRef]
37. Shi, X.; Liu, X.; Xu, C.; Huang, Y.; Chen, F.; Zhu, S. Cross-lingual offensive speech identification with transfer learning for low-resource languages. *Comput. Electr. Eng.* **2022**, *101*, 108005. [CrossRef]
38. Xu, C.; Wang, M.; Ren, Y.; Zhu, S. Enhancing Aspect-based Sentiment Analysis in Tourism Using Large Language Models and Positional Information. *arXiv* **2024**, arXiv:2409.14997.
39. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Chen, X.; Zhang, H.; Tian, X.; Zhu, D.; Tian, H.; Wu, H. Ernie: Enhanced representation through knowledge integration. *arXiv* **2019**, arXiv:1904.09223.
40. González-Carvajal, S.; Garrido-Merchán, E. Comparing bert against traditional machine learning text classification. *arXiv* **2020**, arXiv:2005.13012.
41. Garcia-Mendez, S.; Fernandez-Gavilanes, M.; Juncal-Martinez, J.; Gonzalez-Castano, F.J.; Seara, O.B. Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labelled Corpus. *IEEE Access* **2020**, *8*, 61642–61655. [CrossRef]
42. García, D.; Hu, X.; Rohrer, M. The colour of finance words. *J. Financ. Econ.* **2023**, *147*, 525–549. [CrossRef]
43. Araci, D. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv* **2019**, arXiv:1908.10063.
44. Huang, A.H.; Wang, H.; Yang, Y. FinBERT: A Large Language Model for Extracting Information from Financial Text *. *Contemp. Account. Res.* **2022**, *40*, 806–841. [CrossRef]
45. Shim, H.; Luca, S.; Lowet, D.; Vanrumste, B. Data augmentation and semi-supervised learning for deep neural networks-based text classifier. In Proceedings of the 35th Annual ACM Symposium on Applied Computing, Brno, Czech Republic, 30 March–3 April 2020.
46. Xie, Q.; Dai, Z.; Hovy, E.H.; Luong, M.-T.; Le, Q.V. Unsupervised data augmentation for consistency training. *arXiv* **2019**, arXiv:1904.12848.
47. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
48. Wang, X.; Yang, L.; Wang, D.; Zhen, L. Improved TF-IDF Keyword Extraction Algorithm. *Comput. Sci. Appl.* **2013**, *3*, 64–68.
49. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.
50. Zhang, Q.; Yang, Y. ResT: An Efficient Transformer for Visual Recognition. *arXiv* **2021**, arXiv:2105.13677.
51. Diko, A.; Avola, D.; Cascio, M.; Cinque, L. ReViT: Enhancing vision transformers feature diversity with attention residual connections. *Pattern Recognit.* **2024**, *156*, 110853. [CrossRef]
52. Jawahar, G.; Sagot, B.; Seddah, D. What does bert learn about the structure of language? In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.
53. Ma, X.; Chen, Y.; Liu, X.; Kuang, H. A Chinese short-text oriented intent classification network based on BERT and Bi-GRU. In Proceedings of the 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 26–28 May 2023; pp. 1692–1696.
54. Wu, T.; Wang, M.; Xi, Y.; Zhao, Z. Intent recognition model based on sequential information and sentence features. *Neurocomputing* **2024**, *566*, 127054. [CrossRef]