J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 3603--3621

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on User Intent Recognition Technology Based on Graph Neural Network in Power Marketing Intelligent Customer Service System

Fei Lu[1,✉]

[1] State Grid Dandong Power Supply Company, Dandong, Liaoning, 118000, China

ABSTRACT

How to communicate with users in a timely and effective manner and determine the intentional purpose of customers plays an important role in promoting continuous user interaction and improving service efficiency in the power marketing industry. The article firstly researches on a single-round natural language understanding algorithm based on intent-slot bi-directional interaction, which adopts a bi-directional information flow to realize the bi-directional information interaction between intent and slot. In the intention recognition layer, the interaction attention mechanism is utilized to introduce slot context information. Then the overall design scheme for the construction of an intelligent customer service system for power marketing from dialogue state keeping, multi-round question and answer, model storage to answer visualization is proposed, and the potential functional requirements are analyzed exhaustively. Finally, experiments from various aspects prove the effectiveness of the proposal in this paper. In the comparison experiments on MixATIS with MixSNIPS dataset and DSTC4 dataset, the metrics are improved by 0.3%, 1.5% and 0.5% respectively when comparing GL-GIN model on MixATIS dataset. This leads to the feasibility of the intelligent customer service system for power marketing constructed in this paper.

Keywords: intent recognition, natural language understanding, intelligent customer service system, power marketing

1. Introduction

With the continuous development of social economy, the electric power industry is also growing rapidly, becoming an important support for national economic development. As an important part of supporting the development of the power industry, power marketing is also an indispensable part of the industry [1-3]. Power marketing is an important way for enterprises to understand the actual needs of users, and can communicate effectively with users to improve the economic benefits of

✉ Corresponding author.
E-mail address: lufeiLF123@126.com (F. Lu).
Received 10 June 2024; Revised 25 August 2024; Accepted 20 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-201

enterprises, which is related to the daily life of residents, and in the process of power marketing, the use of power marketing customer service system is particularly important [4-7].

In the context of the new era, China's Internet technology has been further developed, and the problems existing in the traditional power marketing model have gradually come to the fore, unable to meet the business needs of power enterprises [8-9]. Based on this, intelligent electricity service gradually comes into people's daily life, it is in addition to the application of high technology, but also to match more quality customer service, so that customers get a new user experience [10-12]. Through the use of intelligent power marketing system, the service is more three-dimensional, but also a great degree of savings in the cost of operation, for electric energy is a great limit of the use of power, but also reduce the pressure of the staff, thus improving the economic efficiency of the power enterprises, electric power enterprises in order to better development, in order to find the root point in today's society, based on the fierce competition in this era [13-16]. The introduction of user intent recognition technology in the power marketing intelligent customer service system is conducive to helping the system to accurately understand the user's needs and realize the accurate identification of the user's intent, so as to improve the service level of the system and the user's satisfaction [17-19].

The article firstly proposes a method based on the joint modeling of intent recognition and slot filling tasks to realize explicit two-way interaction of information between intent recognition and slot filling in a single-round task-based customer service system, to improve the accuracy rate of the intent recognition and slot filling tasks, and improves the intent recognition task into a multi-intent recognition task, introduces the mechanism of graph attention, and uses the intent labels outputted by the intent recognition task to guide the slot filling The results of the slot filling task are guided by the intent labels output from the intent recognition task. Then the overall requirements of the power marketing intelligent customer service system are analyzed, starting from two perspectives: functional requirements and non-functional requirements for a comprehensive introduction, and overall grasping the two main roles built into the system platform, ordinary users and platform administrators of each core operation and the corresponding key functions. In the experimental part, suitable dialog datasets MixATIS and MixSNIPS are selected for experiments. After verifying the model performance on the datasets, comparative experiments are conducted on single-intent as well as multi-intent datasets. Finally, an overall review of the power marketing intelligent customer service system is conducted.

2. Research on user intention recognition algorithm based on graph neural network

2.1. Overall design idea of the model

In this section, a single-round multi-intent dialog natural language understanding model based on intent-slot bi-directional interaction is proposed and utilized for intent recognition to improve the overall performance of the customer service system and reduce the labor cost. Based on the intent recognition to categorize the intent of the customer's question, the main purpose is to identify the business contained in the question, including complaint and suggestion, business consulting, fault report and repair, and business expansion. According to the identified business categories, the business processing module is substituted into the corresponding words to guide the customer through multiple rounds of dialog to complete the final business processing. The model mainly consists of word embedding coding layer, intent recognition layer based on slot-intent interaction and slot filling layer based on intent-slot interaction [20]. The model in this section is named as RCGT.The RCGT model architecture is shown in Fig. 1.

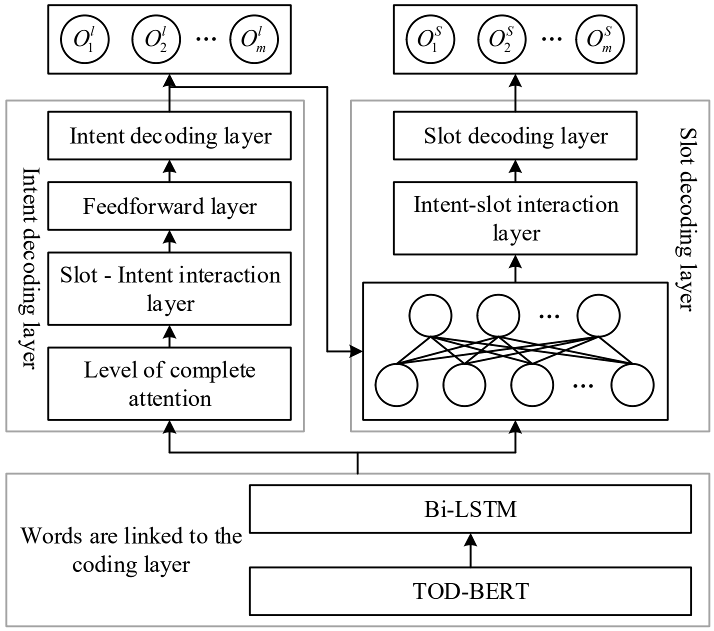


Fig. 1. Architecture of the RCGT model

2.2. Construction of the RCGT model

2.2.1. Word Embedding Coding Layer:

1) TOD-BERT pre-training layer. The structure of TOD-BERT is shown in Fig. 2.TOD-BERT is a variant of BERT designed specifically for task-oriented customer service systems. The powerful language comprehension capabilities of BERT are utilized and adapted and optimized to better handle the tasks of interacting with users, understanding their intentions, and generating appropriate responses.

For a sentence $U=\{u_1,u_2,\ldots,u_n\}$, where, n denotes the number of words after the sentence has been split. Since TOD-BERT adds [CLS] to the beginning of the sentence and [SEP] tags to the end of the sentence at the encoding stage, and [USR] and [SYS] tags are also added in multiple rounds of the dialog, the resulting sequence is $X=\{x_1,x_2,\ldots,x_n\}$ after TOD-BERT encoding.

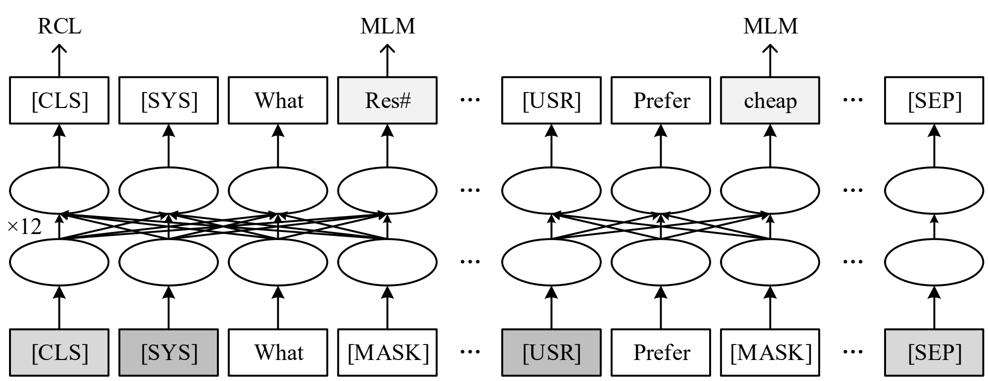


Fig. 2. TOD-BERT structure

2) Bi-LSTM coding layer. After pre-training coding, this subsection introduces the Bi-LSTM coding module, which serves to make full use of the temporal features in the sentence to capture the dependencies over longer distances in the sentence and model the contextual information of the sentence.The schematic diagram of Bi-LSTM is shown in Fig. 3.

Since the LSTM module itself is unidirectional, it can only model the contextual relationship from left to right but cannot encode the information from right to left, when encountering sentences with inverted structure, the modifying components often appear after the modified words, the problem of dependency between the front and back of words can be solved by adopting the Bi-LSTM.The Bi-LSTM consists of two LSTM layers, for a sequence $X = \{x_1, x_2, ..., x_n\}$. Bi-LSTM is capable of bi-directional encoding from left to right and from right to left, and finally the hidden states in both directions are spliced together to get the hidden state $H = \{h_1, h_2, ..., h_n\}$ which contains the context information.

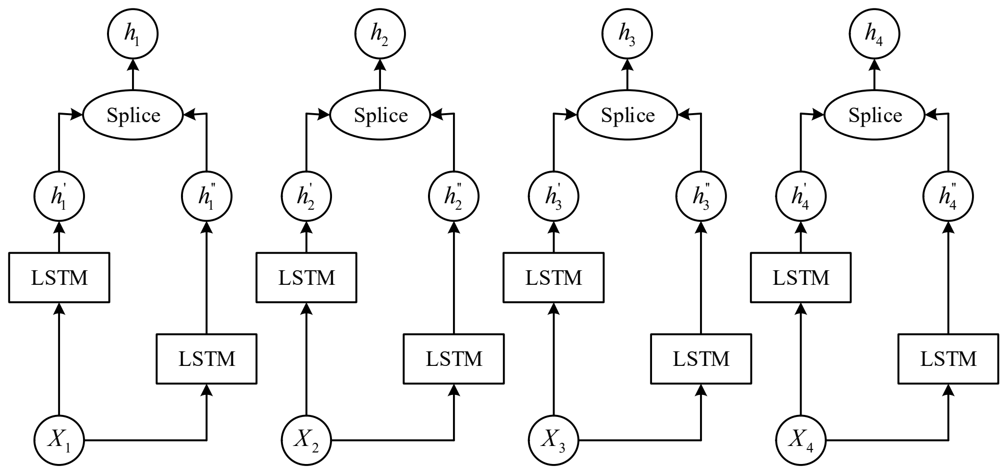


Fig. 3. Schematic diagram of the Bi-LSTM

The hidden state h_i and cell state c_i of the LSTM model at moment i are determined by the input gate $input_i$, forgetting gate $forget_i$ and output gate $output_i$ computed from the current word embedding $x_i \in X$, the cell's previous moment state c_{i-1}, the hidden state h_{t-1} of the previous moment, and the cell's temporary state $temp_t$. The computational procedure of the LSTM is shown in Eqs. (1) to (6).

$$forget_i = sigmoid(W_f[h_{i-1}, x_i] + b_f) \tag{1}$$

$$input_i = sigmoid(W_i[h_{i-1}, x_i] + b_i) \tag{2}$$

$$output_i = sigmoid(W_o[h_{i-1}, x_i] + b_o) \tag{3}$$

$$temp_i = \tanh(W_t[h_{i-1}, x_i] + b_t) \tag{4}$$

$$c_i = forget_i \odot c_{i-1} + input_i \odot temp_i \tag{5}$$

$$h_i = output_i \odot \tanh(c_i) \tag{6}$$

2.2.2. Intent Recognition Layer Based on Slot-Intent Interaction:

1) Full Attention Layer. The full attention layer in this subsection employs an improved attention mechanism that allows the model to more accurately focus on words that are closely related to a specific label by incorporating label information directly into the attention computation process, while performing intent recognition and slot filling tasks. The implementation of the full attention mechanism is shown in Eqs. (7) to (8).

$$\alpha_{fal} = softmax(H_{BiLSTM} W^{fcl}) \tag{7}$$

$$\alpha_{faS} = softmax(H_{BiLSTM} W^{feS}) \tag{8}$$

where α_{faI} and α_{faS} denote the attention coefficients of the two tasks of intent recognition and slot filling, respectively, reflecting the degree of relevance of the intent and slot labels to the hidden state $H_{BiLSTM} = \{h_1, h_2, \ldots, h_n\}$, and $W^{feI} \in \mathbb{R}^{dx|I^L|}$, $W^{fcS} \in \mathbb{R}^{dx[S^L]}$, I^L, and S^L denote the number of labels for the intent recognition and slot filling tasks, respectively.

Contextual representations H_{faI} and H_{fas} containing the intent and slot labeling information are finally obtained, as shown in Equations (9) to (10).

$$H_{faI} = H_{BiLSTM} + \alpha W^{fcI} \tag{9}$$

$$H_{faS} = H_{BiLSTM} + \alpha W^{fcS} \tag{10}$$

2) Slot-intention interaction layer. This subsection adopts an improved slot-intention interaction attention mechanism. Firstly, the hidden states are mapped to three different spaces through a linear transformation to obtain the query vector Q_I, key vector K_S and value vector V_S needed for slot-intention interaction attention, which are realized as shown in Eqs. (11) to (13).

$$Q_I = W_{qI} H_{faI} \tag{11}$$

$$K_s = W_{kS} H_{faS} \tag{12}$$

$$V_s = V_{kS} H_{faS} \tag{13}$$

where $W_{ql} \in \mathbb{R}^{D_q \times D_H}, W_{ks} \in \mathbb{R}^{D_t \times D_M}, W_{vs} \in \mathbb{R}^{D_t \times D_M}$ denotes the trainable weight matrix

After that, Q_I is used as the query vector, and K_S and V_S are used as the key vector and value vector to do the attention operation, as shown in Equation (14).

$$H_{S2I} = softmax(\frac{Q_I K_S^T}{\sqrt{d_k}}) V_s \tag{14}$$

where d_k is the vector dimension of the K_S matrix.

Up to this point, the model explicitly obtains a contextual representation of the degree of association of the slot representation to the intent, similar to the native Transformer, and performs residual concatenation and layer normalization on the output, which ultimately yields a contextual representation of the intent after incorporating semantic information about the slots. As shown in equation (15).

$$H_{S2LAtm} = LayerNorm(H_{S2I} + H_{faS}) \tag{15}$$

3) Feedforward layer. The feedforward layer (FFN) generally consists of two fully connected layers and an activation function. The working principle of the feedforward layer is shown in Equation (16).

$$FFN(x) = W_2(LeakyReLU(W_1 x + b_1)) + b_2 \tag{16}$$

where W_1, W_2 is the learnable weight matrix, b_1, b_2 is the learnable bias, and x denotes the input to the FFN.

The hidden representation is then normalized by one more residual join and layer normalization, which further makes LeakyReLU work better, and the process is shown in Equation (17).

$$H_{FFN} = LayerNorm(FFN(H_{S2LAtn}) + H_{S2LAtn}) \tag{17}$$

4) Intention decoding layer. The intention recognition module based on slot-intention interaction firstly goes through TOD-BERT+BiLSTM encoding to get the hidden state $H_{BiLSTM} = \{h_1, h_2, \ldots, h_n\}$ of the sentence, and after fusion and enhancement of the intention and slot information through the full-attention layer, slot-intention interaction layer, and feed-forward layer, it gets the contextual representation H_{FFN}, and finally, in this subsection, it realizes the intention decoding.

For the multi-intent recognition task, Sigmoid is used as the activation function, and the hidden state is passed through the Sigmoid function to get the probability of each intent label after the fully connected layer. The specific methods are shown in Eqs. (18) to (20).

$$Y^I = Sigmoid(W_{fcI} H_{FFN} + b_{fcI}) \tag{18}$$

$$Y^I = \{y_1^I, y_2^I, \ldots, y_n^I\} \tag{19}$$

$$O^I = \{o_1^I, o_2^I, \ldots, o_m^I\} = \{Y^I \mid y_i^I > t^I\} \tag{20}$$

2.2.3. Slot Filling Layer Based on Intent-Slot Interaction:

1) Multi-Intent-Slot Bitmap Interaction Layer. GAT is a graph neural network based on the attention mechanism, which achieves the learning of node representations and relationships between nodes by assigning different weights to the neighboring nodes of each node. First, for intent label encoding, this subsection employs GloVe as an external knowledge base for initializing intent label embeddings [21]. The intention label embedding E^I and the hidden state H_{BiLSTM} obtained from the word embedding encoding layer are spliced to obtain the input $H_v = \{h_{v1}, h_{v2}, \ldots, h_{vk}\} = \{h_1, h_2, \ldots, h_n, e_1^I, e_2^I, \ldots, e_m^I\}$ of GAT, where $h_{vi} \in \mathbb{R}^F$.

The specific process of GAT adopted in this subsection is described next. Similar to the attention mechanism, GAT does the computation of the attention coefficients for each neighbor node N_i of node i on itself, as shown in Equation (21).

$$e_{ij} = LeakyReLU([W_g h_{vi}, W_g h_{vj}] a^T) \tag{21}$$

where LeakyReLU is the nonlinear activation function. $a \in \mathbb{R}^{2F'}$ and $W_g \in \mathbb{R}^{F \times F}$ are trainable weight matrices, and $[\cdot]$ denotes the splicing operation. Next, after the activation function Softmax, the attention value e_i of each neighbor node N_i of v_i is normalized to obtain the attention weight α_{ij} as shown in Equation (22).

$$\alpha_{ij} = softmax_j(e_{ij}) \tag{22}$$

After that, the aggregation of the features of each neighboring node is achieved through the attention weights and the attention coefficients are obtained as shown in Equation (23).

$$h_i' = elu(\sum_{j \in N_i} \alpha_{ij} W_g h_{vj}) \tag{23}$$

where elu is the nonlinear activation function. The computational flow of the multiple attention mechanism is shown in Equation (24).

$$h_i' = concat_k(elu(\sum_{j \in N_i} \alpha_{ij}^k W_g^k h_{vj})) \tag{24}$$

Exceptionally, for the last layer, the mean value operation is used instead of the splicing operation, as shown in Eq. (25).

$$h_i^{'} = elu(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i}\alpha_{ij}^{k}W_s^{k}h_{vj}) \tag{25}$$

After l layer of graph attention layer computation, the output h_{GAT} of the last layer of graph attention is obtained as shown in Equation (26).

$$H_{GAT} = \{h_1^l, h_2^l, \ldots, h_n^l\} \tag{26}$$

2) Slot Decoding Layer. This subsection employs the Conditional Random Field (CRF) model to model the correlation between adjacent lexical labels. The output H_{GAT} of the intent-slot interaction layer is mapped to the label space by a linear transformation, as shown in Equation (27).

$$H_{fcS} = W_{fcS}H_{GAT} + b_{fcS} \tag{27}$$

where $H_{\rho s} \in \square^{n\times k}$. k denotes the number of slot labels. The weight matrix W_{fes} and bias $b_{\rho as}$ are trainable parameters.

The conditional probability distribution of CRF is shown in Eq. (28) using maximum group factorization.

$$P(Y_s \mid H_{feS}) = \frac{1}{Z(H_{feS})}\exp(\sum_{i=1}^{n}F(y_{i-1}, y_i, H_{feS})) \tag{28}$$

where Y_S denotes the slot label prediction result, $Z(H_{fcs})$ denotes the normalization factor, and $F(y_{i-1}, y_i, H_{fcS})$ is the scoring function, which denotes a score from y_{i-1} to y_i. The scoring function is calculated as shown in Equation (29).

$$F(y_{i-1}, y_i, H_{feS}) = \sum_{k=1}^{K}\lambda_k t_k(y_{i-1}, y_i, H_{feS}) + \sum_{l=1}^{L}\mu_l s_l(y_i, H_{feS}) \tag{29}$$

The sum of all possible paths is calculated by the normalization factor, which is shown in Equation (30).

$$Z(H_{feS}) = \sum_{y=1}^{n}\exp(\sum_{k=1}^{K}\lambda_k t_k(y_{i-1}, y_i, H_{feS}) + \sum_{l=1}^{L}\mu_l s_l(y_i, H_{feS})) \tag{30}$$

Finally, the Viterbi algorithm is used to efficiently find the most probable labeling sequence Y_5 as shown in Equation (31).

$$O_s = argmax_{Y_s} P(Y_s \mid H_{feS}) \tag{31}$$

2.2.4. Loss function. In this paper, the ASL loss function is used as the loss function for the intention recognition task, and ASL mainly realizes the downweighting of negative simple samples in the following two ways.

1) For the negative simple samples with very low prediction probability, a threshold t^m is introduced, which will be overall, and the negative simple samples with probability less than t^m are directly discarded, thus further reducing the weight of the simple negative samples. As shown in equation (32).

$$o_i^{shiftI} = \max(o_i^I - t^m, 0) \tag{32}$$

where the threshold t^m is a hyperparameter, and t^m is set to 0.2 in this paper.

2) Asymmetric attention is applied to the samples, learning weight parameter γ_+ is applied to the positive samples, and learning weight parameter γ_- is applied to the negative samples, by decoupling

the learning weight parameter and setting the learning weight parameter $\gamma_- > \gamma_+$, so as to realize the question of downweighting the negative simple samples, and maintain the attention to the positive samples. The calculation of loss function for positive and negative samples is shown in equations (33) and (34).

$$\mathrm{L}_+ = (1 - o_i^r)^{\gamma_+} \log(o_i^r) \tag{33}$$

$$\mathrm{L}_- = (o_i^{shiftI})^{\gamma_-} \log(1 - o_i^{shiftI}) \tag{34}$$

where γ_+ and γ_- are hyperparameters greater than or equal to 0. In this subsection, γ_+ is set to 0 and γ_- is set to 4.

The loss function for the intention recognition task is shown in Equation (35).

$$\mathrm{L}_I = \sum_{i=1}^{N_r} (-\hat{y}_i \mathrm{L}_+ - (1 - \hat{y}_i)\mathrm{L}_-) \tag{35}$$

For the slot filling task, the negative log likelihood (NLL) is used as the loss function as shown in Equation (36).

$$\mathrm{L}_s = -\log P(Y_s \mid H_{fcS}) \tag{36}$$

In order to achieve the joint optimization of the two tasks, the loss functions of the two tasks are weighted and summed to obtain the final loss function as shown in Equation (37).

$$\mathrm{L} = \mathrm{L}_I + \beta \mathrm{L}_S \tag{37}$$

3. Demand analysis of intelligent customer service system for power marketing

The main work of this paper is to design and implement an intelligent customer service system for power marketing, which requires the system to first accept the user's question text, analyze the intent and slot information related to the question and answer in the sentence using a graph neural network based user intent recognition algorithm, and integrate it into the user's conversation state object. Each user's conversation state object is unique and stored in the cache Redis, with a fixed duration renewal period for conversation retention and conversation state update. Using the intent slots of the model, the system can answer questions by querying the already pre-built knowledge graph of the power domain, and the system embellishes the answers and returns them to the user side. The process knowledge graph is stored in the graph database Neo4j, while the relational data is stored in Mysql [22]. The Q&A history of the user and the system in the process of Q&A is stored in a fixed directory on the server and is done at regular daily intervals using a timed task configuration module developed by the system and is archived using an object storage service, which is chosen to be stored in Minio in the current version. This section focuses on the functional and non-functional requirements involved in the platform of the power marketing intelligent customer service system, and will divide the users while giving the roles of the platform and show the corresponding use cases under different user perspectives. The system use case diagram is shown in Figure 4.

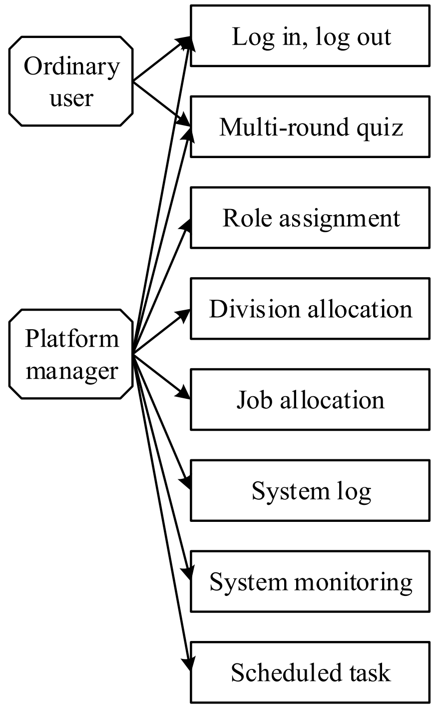


Fig. 4. System use case diagram

3.1. Analysis of platform user roles

Since the creation of this electric power marketing intelligent customer service system platform, it has two built-in roles, which are the ordinary user of the platform and the corresponding administrator of the platform. The administrator has the authority to create and assign roles, and the platform allows the administrator to make further fine-grained division of roles according to the needs, and the roles with different permissions can be found and blocked directly through the display and hiding of the browser menu interface. These roles are customized roles for easy expansion, and the use cases discussed after this section are based on the user built-in roles as an example.

3.1.1. General user roles. Ordinary users are the users who access the basic Q&A function of the power marketing intelligent customer service system platform, the user directly inputs the text of the question in the browser, and according to the answer displayed in the output box system can be continued to pursue the question and the question can be changed [23]. The process of using the Q&A service, the service provided by the system to the user involves the cooperation between multiple module microservices, while it is transparent and senseless to the user.

3.1.2. Platform administrator role. In the power marketing intelligent customer service system, the platform administrator has the ability to be responsible for the operation of the system, in addition to the ability to use ordinary user roles in the system. Administrators can customize the creation of role types to fulfill the needs of more customers, and divide the authority granularity more finely. In terms of monitoring system performance, administrators have access to system logs in a way that matches the concepts often described in monitoring and link tracing in distributed systems.

3.2. System requirements analysis

3.2.1. Functional requirements:

1) User login and logout. According to the previous analysis of the platform user roles, it can be seen that the platform provides two built-in user roles, corresponding to ordinary users and platform

administrators. The platform needs to differentiate authentication according to the different roles that users belong to, and the function entrances displayed on the interface will also be different at that time. Meanwhile, in order to provide higher security during the login process, the back-end user service introduces a randomly generated dynamic verification code, which needs to be provided together.

2) Management of multiple rounds of user questions and answers. Functional users submit their questions to the system backend through the question input box provided in the front-end of the home browser. The system then tries to take out the user's dialog state from the cache, and re-create a blank state for the user if it has expired, after which a remote call is made to the packaged model algorithm service, and the model recognizes the semantics embedded in the user's text. Subsequently, after returning the results, the mainstream process in the system backend needs to complete the tracking of the current dialog state, querying the knowledge graph stored in the graph database for the dialog task that meets the conditions, and then obtaining the results also needs to complete the updating of the dialog state, the storage keeping of the dialog state, and the document recording of the current dialog, in order to wait for the execution of the archiving task configured in the timed task.

3) Platform administrator timed task management. Multi-round Q&A module needs to rely on timed tasks to archive the information of Q&A records. Timed task management belongs to the indispensable basic services of multi-round Q&A service, and it can also be extended by the administrator to configure timed tasks for the platform. At present, the timed task management function belongs to the scope of system administrator's authority, and only the platform administrator has the authority to operate this function. In this function, the role of the operation is allowed to query the timed tasks that have been run in the current platform, and it supports locking according to multiple combination conditions.

4) System management function of platform administrator. The current system management function involves the addition, editing, deactivation and deletion of users, roles, departments, positions and other functional modules, which belongs to the scope of the system administrator's authority, and only the platform administrator and the customized roles may have the authority to operate such functions. In this function, the role of operation is allowed to query the data of each module in the current platform, and it supports locking according to multiple combination conditions.

5) Platform administrator system monitoring function. The log management function belongs to the scope of authority of the system administrator, and only the platform administrator and the customized roles may have the operation authority of this function. The operation log plays an important role in checking the operation status of the system and dealing with system service exceptions. At present, the platform already supports the export of operation logs to Excel table files.

3.2.2. Non-functional requirements:

1) Platform system operation easy to use. Power marketing intelligent customer service system platform to be user-oriented simple and easy to use, interface function operability to ensure a good user experience.

2) Get the platform response in real time. In the process of users using the platform for multiple rounds of questions and answers, the design of multiple micro-service modules and algorithmic modeling services, knowledge map access to cooperate with each other, a question and answer delay needs to be controlled in the user experience can be accepted within the scope.

3) Sustainable expansion of platform functions. When the demand side of the platform has new functional requirements for integration and access, the system expansion should be simple and

convenient enough. Since the Spring Cloud Alibaba microservice architecture has been adopted, the modularized development of the back-end interfaces follows the corresponding Controller, Service, Dao, and Bean layers, and follows the high cohesion and low-coupling design principle of software development. The design principles of software development are followed.

4) Sustainable maintenance of platform system. During the development of the system platform, all mainstream frameworks are used in maintenance. The model algorithm service is packaged in Django as an HTTP service for network invocation, and the front-end uses Vue 2 with Element-Ui component library for page development.

4. Experimental results and analysis

4.1. Data preparation

In this paper, the effectiveness of the system is verified using the power marketing intelligent customer service system dialog datasets MixATIS and MixSNIPS, which mainly have about one million multi-round dialogs as well as more than twenty million single sentences. In addition, the number of rounds per conversation ranges from 3 to 85, with an average of 20 rounds, and each single sentence has an average of 7.6 words. The main statistical features of the dataset are shown in Table 1.

Table 1. Statistical characteristics of data collectors

Attribute	Numerical value
Total dialogue	1,023,251
Total number of sentences	20,443,329
Total word	150,715,202
Mean word	7.6
The average number of rounds	23
Maximum wheel number	85
Minimum wheel number	3

The number of samples corresponding to different number of dialog rounds is shown in Figure 5. Due to space limitation, this paper only shows conversations with less than 50 rounds. We can see that most of the rounds are between 9 and 30, while the largest proportion of rounds is 14 rounds. This suggests that long-term dependency between contexts is a distinctive feature of the MixATIS and MixSNIPS datasets.

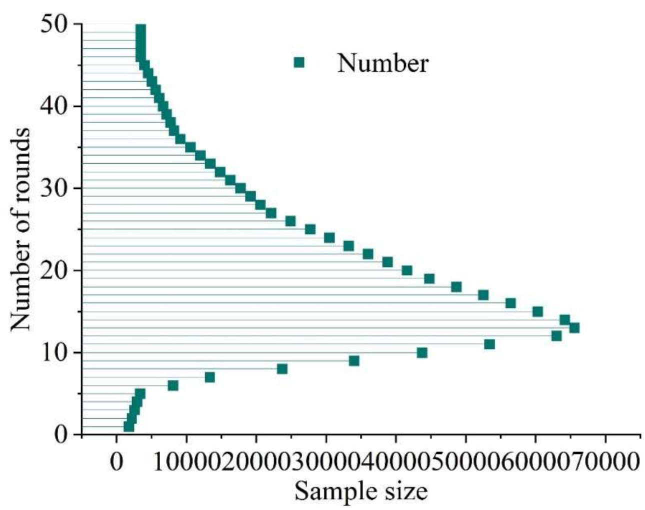


Fig. 5. The number of different dialog wheel Numbers corresponds to the sample number

4.2. *Comparison of Algorithms*

In order to better evaluate the RCGT, the experiments compare the Baseline model with the state-of-the-art model, and in the subsequent experiments, the experimental parameters of the comparison model as well as the final modelability are taken from the best experimental results in the original paper of the Baseline model. A brief description of the models used for comparison is given below:

Attention BiRNN: A network model for the joint task of intent recognition and slot filling is proposed, which is an alignment model based on the Attention mechanism.

Slot-Gated: Adopts a “gate” mechanism to control the preservation of feature information during the model iteration. The model focuses on learning the feature information relationship between the two tasks and obtaining better feature information through global optimization.

Bi-Model: A bimodel-based recurrent neural network parsing network structure is proposed, and two bi-directional long and short-term memory network pieces are utilized to compute the interaction between two tasks.

SF-ID: An introduction of SF-ID network is proposed to enhance the two-task bi-directional interactions to achieve mutual learning between the two tasks and improve the performance.

Stack-Propagation: In order to better integrate the intent information and then further guide the slot-filling task, the model can directly use the intent feature information as the input to the slot-filling task, which ultimately realizes capturing the intent semantic feature information.

Joint Multiple ID-SF: A multitasking framework utilizing the Slot-Gate mechanism is proposed for solving the joint modeling problem.

AGIF: A word-level Adaptive Graph Interaction Network is proposed, which is used to establish word-to-word intent interactions, thus providing fine-grained intent information to guide the slot-filling task.

GL-GIN: This model is a fast unnatural regression model that utilizes a graph attention network to coordinate feature information for both tasks and has a respectable slot decoding speed.

The results of the comparison experiments on the MixATIS vs. MixSNIPS dataset and the DSTC4 dataset are shown in Tables 2 and 3. In this paper, we use F1 score to evaluate the performance of slot filling task, accuracy to evaluate the intention recognition task, and use overall accuracy to evaluate the performance of utterance-level semantic parsing, and these metrics are improved by 0.3%, 1.5% & 0.5% respectively when comparing the GL-GIN model on the MixATIS dataset.

Table 2. Comparison results of MixATIS and MixSNIPS data sets

Model	MixATIS(%)			MixSNIPS(%)		
	Overall	Slot F1	Intent Acc	Overall	Slot F1	Intent Acc
Attention BiRNN	37.7	87.2	74.2	58	90.4	95.1
Slot-Gate	36.6	86.8	63.2	56.8	87.9	93.8
Bi-Model	35.1	84.1	69.2	54.1	87.3	93.3
SF-ID	34.9	88.2	64.7	57.9	90.8	95.7
Stack-Propagation	42.2	87.8	72.7	72.6	92.8	96.3
Joint Multiple ID-SF	35.7	83.2	75	63.2	89.6	96.4
AGIF	39.7	86	72.9	75.8	93.6	97.5
GL-GIN	43.8	87	76.6	75.8	95.7	95.1
RCGT	44.1	88.5	77.1	75.5	93.6	95.5

Table 3. Compare experimental results on dstc4 data set

Model	DSTC4		
	Overall Acc(%)	Slot F1(%)	Intent Acc(%)
Attention BiRNN	34.2	44.9	43.1
Slot-Gate	32.3	43.3	43.2
Bi-Model	32.3	44.1	41.4
SF-ID	32.9	52.9	41.4
Stack-Propagation(concatenation)	34.2	52.5	45.1
Stack-Propagation(sigmoid-decoder)	31.5	49.4	38.9
Joint Multiple ID-SF	30.4	48.8	37.6
AGIF	34.6	54.7	46.3
GL-GIN	34.4	50.7	44
RCGT	35.8	54.7	48.8

In addition to this, this section also explores the performance of RCGT on single-intent datasets, and the results of comparison experiments on ATIS as well as SNIPS datasets are shown in Table 4. From the results of the tables, it can be seen that the RCGT proposed in this section achieves the best F1 score, recall, and precision on single-intent as well as multi-intent datasets, which indicates that the RCGT model achieves the state-of-the-art model performance.

Table 4. The comparison of the results of the ATIS and SNIPS data sets

Model	ATIS(%)			SNIPS(%)		
	Overall	Slot F1	Intent Acc	Overall Acc	Slot F1	Intent Acc
SF-ID	86.2	97.7	95	79.2	93.4	97.7
Stack-Propagation	86.5	94.6	97.1	85	94.8	97.3
Joint Multiple ID-SF	--	93.7	95.7	--	88.4	96.8
AGIF	87.6	95.5	95.7	87.5	94.9	97.6
GL-GIN	86.2	98	95.3	86.7	95.2	98.9
RCGT	88	96.7	97.6	89.8	96.6	98.9

In terms of model training, this paper records the convergence of RCGT, GL-GIN, AGIF, and SF-ID in the experiments, and the changes of slot-filling F1 and intent-recognition Acc for the four models on the MixATIS dataset are shown in Figs. 6 and 7.The changes of slot-filling F1 and intent-recognition Acc for the four models on the MixSNIPS dataset are shown in Figs. 8 and 9. From the figures, it can be seen that the RCGT model proposed in this paper is able to reach convergence in fewer training batches during the training process, which again proves the effectiveness of the model proposed in this paper in terms of model convergence. From the comparative experimental results, it can be seen that using the multi-task graphical attention network to draw the "intent-slot" interaction graph, and making the two tasks guide each other, compared with the past studies that simply use the intent recognition feature information to guide the slot filling task, the proposed way of utilizing the two types of information to guide each other has a better effect, and also proves that the proposed RCGT model can achieve convergence in fewer training batches during the training process. It also proves that the RCGT proposed in this paper can sufficiently interact the information of the two tasks so that they can guide each other and improve the classification accuracy.

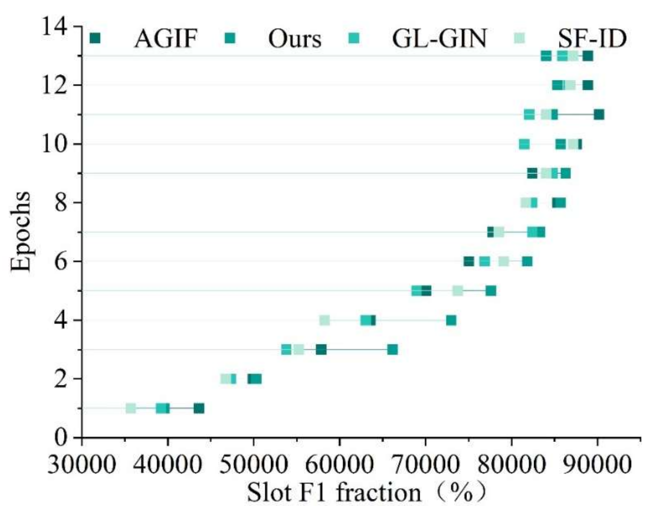


Fig. 6. In the MixATIS data set, the four model slots populate F1 changes

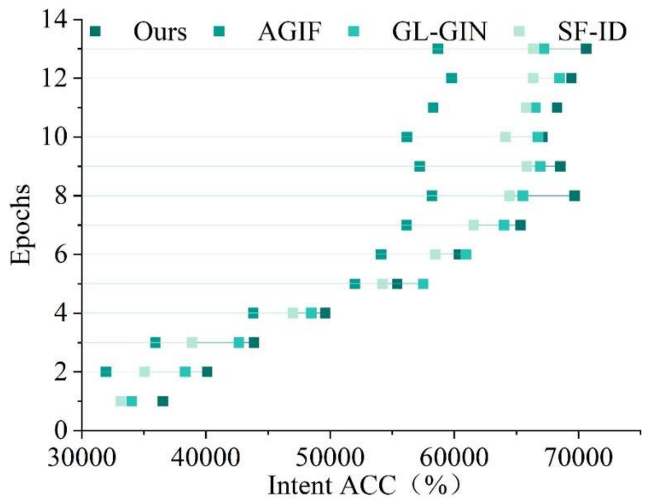


Fig. 7. Four types of patterns of the MixATIS data set identify the acc changes

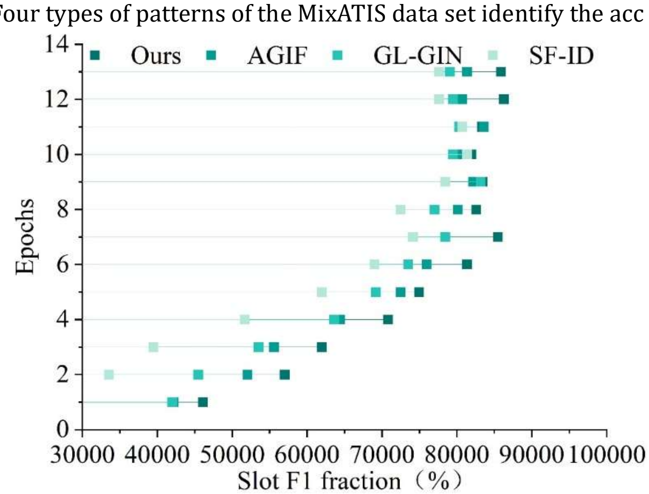


Fig. 8. In the MixSNIPS data set, the four model slots populate F1 changes

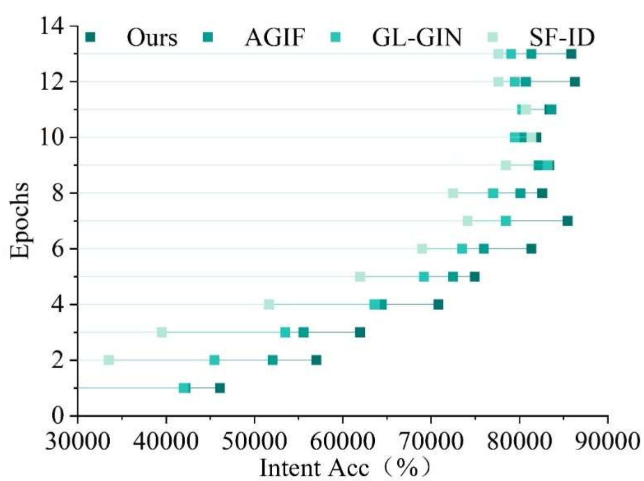


Fig. 9. Four types of patterns were intended to identify acc changes

4.3. Ablation experiments

The comparison of inference rates is shown in Table 5. From the table, it can be seen that the RCGT model proposed in this paper has relatively less inference delay time in the decoding slot when reasoning on the test set of MixATIS using a batch size of 32. This is because the global pointer model decoding slot avoids the inference delay caused by the verbatim decoding of the regression model and also avoids the time spent on the computation of the softmax function of the non-regression model. It can be seen that the RCGT model proposed in this paper has improved performance and efficiency with a decoding delay of 5.1s.

Table 5. Speedup Comparison

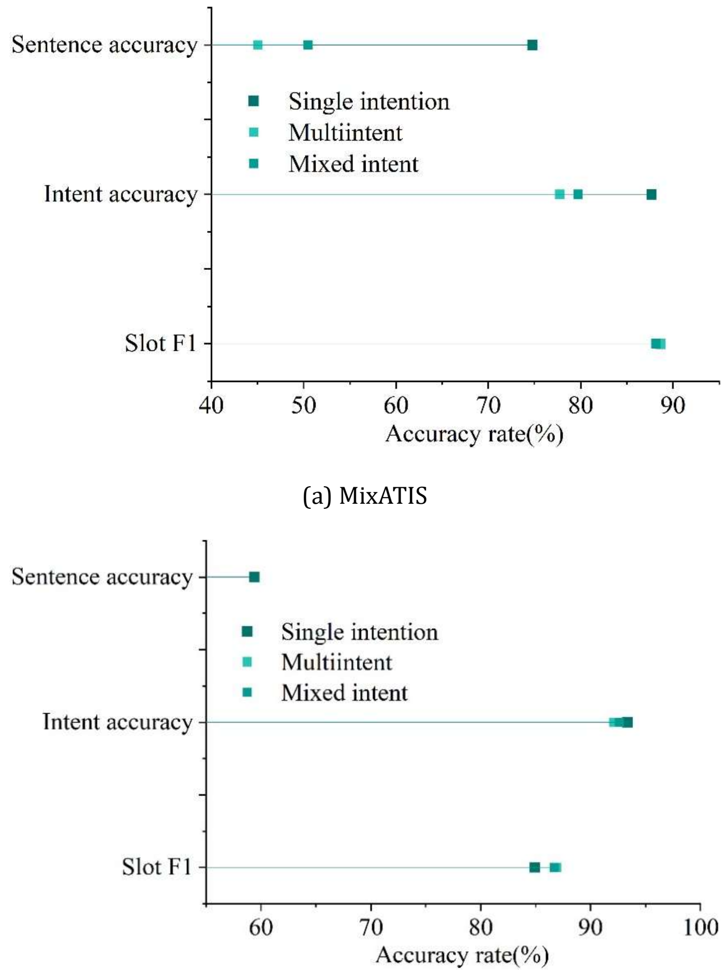
Model	Decode Latency(s)	Speedup
Stack-propagation	33	2.5x
Joint Multiple ID-SF	44.5	2.1x
AGIF	46.7	0.5x
GL-GIN	3	11.8x
Ours	5.1	14.2x

The ablation experiments are shown in Table 6. From the table, it can be seen that the RCGT model proposed in this chapter outperforms the MSLN-GP model in several metrics, which is due to the fact that the multidimensional type-slot labeling interaction network is capable of filtering out useful slot information, which can help in the subsequent decoding of the intent, and at the same time, it is able to enhance the implicit correlation between intents and slots, so that the information can be obtained from each other to complement each other.

Table 6. Ablation experiment

Model	MixATIS			MixSNIPS		
	Slot F1	Intent Acc	Overall Acc	Slot F1	Intent Acc	Overall Acc
w/o TOD-BERT	85.6	86.6	42.6	93	97.6	82.6
w/o Bi-LSTM	87.4	75.6	46.7	94	97.4	83.1
w/o Full-Attn	89.1	78.9	47.6	97.2	97.4	82.5
w/o S2I-CoAttn	88.6	79.4	48.9	98.1	97.7	83.9
w/o I2S-Inter	86.9	80.3	48.9	97.5	98.3	82.7
RCGT	88.5	78.3	49.5	98.7	99	83.5

In order to further analyze the effect of the proposed model under single and multiple intentions, the test set is divided into single and multiple intentions based on the number of intentions in the discourse.The experimental results for single and multiple intentions are shown in Figure 10 (Figure a shows the experimental results on the MixATIS test set, and Figure b shows the experimental results on MixSNIPS).The single and multiple intentions for MixATIS test set samples are 145 and 682, respectively, and the single-intent and multi-intent samples are 455 and 1753 for the MixSNIPS test set.From the results, it can be seen that:(1) the single-intent and multi-intent samples are basically the same in slot F1, which indicates that the proposed model well utilizes global pointers to solve the slot incoherence problem. (2) In terms of intent accuracy, compared with the MixSNIPS dataset, the gap between the MixATIS dataset in the two samples is relatively large, which may be due to the imbalanced distribution of intent labels in MixATIS, with only 17 intents in the training set and 18 intents in the test set. Overall, the proposed model can handle both single-intent and multi-intent samples well. (3) As far as sentence accuracy is concerned, it can be seen that single-intent samples work better than multi-intent samples, due to the fact that the single-intent task is simpler than the multi-intent one, which is realistic. To summarize, the model proposed in this chapter can obtain good results in both single-intention and multi-intention samples.



(a) MixATIS

(b) MixSNIPS

Fig. 10. Experimental results of single-intent and multi-intent

4.4. Overall system evaluation

Five participants were recruited for the experiment and evaluated using the evaluation method. In the experiment, participants were asked to score the fluency, information and system satisfaction of the power marketing intelligent customer service system established in this paper, with the score ranging from 0 to 5, and the larger the score, the better. In order to eliminate personal bias, each participant was tested with 50 sets of conversations separately to ensure the accuracy of the evaluation. The evaluation of the power marketing intelligent customer service system is shown in Table 7, and the system scores high in both fluency and satisfaction, with scores of 8.48 and 8.8, respectively.It shows that the customer service system constructed in this paper makes full use of the domain data, and provides users with rich information.

Table 7. Evaluation of intelligent customer service system of electric power marketing

	Fluency	Information content	Satisfaction
Participant 1	8.8	7.2	8.4
Participant 2	9.2	8.3	8.8
Participant 3	8.7	7.9	9.6
Participant 4	8.3	8.1	8.4
Participant 5	7.4	8.5	8.8
Average score	8.48	8	8.8

5. Conclusion

This paper mainly focuses on the in-depth analysis and research of natural language understanding algorithms in the power marketing intelligent customer service system, and conducts role analysis as well as system requirement analysis for platform users. The article draws the following conclusions:

The RCGT model is tested on MixATIS and MixSNIPS datasets, and the results show that in this paper's model on MixATIS dataset compared with the GL-GIN model, the values of Slot F1 and Intent Acc are improved by 1.5% and 0.5% respectively

In the reasoning rate comparison experiment, the RCGT model proposed in this paper is improved in performance and efficiency with an acceleration of 14.2x, thus verifying the effectiveness of the natural language understanding algorithm proposed in this paper in the power marketing intelligent customer service system.

References

[1] Newbery, D. M., Reiner, D. M., & Ritz, R. A. (2019). The political economy of a carbon price floor for power generation. The Energy Journal, 40(1), 1-24.

[2] Hu, J., Harmsen, R., Crijns-Graus, W., Worrell, E., & van den Broek, M. (2018). Identifying barriers to large-scale integration of variable renewable electricity into the electricity market: A literature review of market design. Renewable and sustainable energy reviews, 81, 2181-2195.

[3] Baroche, T., Pinson, P., Latimier, R. L. G., & Ahmed, H. B. (2019). Exogenous cost allocation in peer-to-peer electricity markets. IEEE Transactions on Power Systems, 34(4), 2553-2564.

[4] Sun, C., Zang, Y., Sun, J., & Liu, Y. (2019, May). Research on Lean Management of Power Marketing Strategy in Power Grid Enterprises. In 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia) (pp. 3091-3094). IEEE.

[5] Tetiana, S., Tetiana, K., Iryna, S., Oleksandr, T., Mykola, S., & Yevgeniy, H. (2021). Estimation of management effectiveness of electricity supply enterprises in emerging economies. TEM Journal, 10(1), 238.

[6] Ajibola, A. A., Sodeinde, G. M., Aderemi, T. A., & Yusuf, M. O. (2021). Impact of Electricity Supply on the Performance of Small and Medium-Scale Enterprises (SMEs) in Nigeria: A Case Study. Economic Insights-Trends & Challenges, (4).

[7] Joseph, A., & Balachandra, P. (2020). Energy internet, the future electricity system: Overview, concept, model structure, and mechanism. Energies, 13(16), 4242.

[8] Chen, T., Pourbabak, H., & Su, W. (2019). Electricity market reform. In The energy internet (pp. 97-121). Woodhead Publishing.

[9] Gao, B., Han, X., Ye, F., & Li, Y. (2023). Internet usage and household electricity consumption. Frontiers in Energy Research, 11, 1323077.

[10] Xiong, J., Ren, J., Chen, L., Yao, Z., Lin, M., Wu, D., & Niu, B. (2018). Enhancing privacy and availability for data clustering in intelligent electrical service of IoT. IEEE Internet of Things Journal, 6(2), 1530-1540.

[11] Lukić, J., Radenković, M., Despotović-Zrakić, M., Labus, A., & Bogdanović, Z. (2017). Supply chain intelligence for electricity markets: A smart grid perspective. Information Systems Frontiers, 19, 91-107.

[12] Marinakis, V., Doukas, H., Tsapelas, J., Mouzakitis, S., Sicilia, Á., Madrazo, L., & Sgouridis, S. (2020). From big data to smart energy services: An application for intelligent energy management. Future Generation Computer Systems, 110, 572-586.

[13] Sun, S., Gu, H., Peng, C., Zhou, H., & Seng, D. (2022). The Architecture of an Intelligent Technical Support System for Electricity Spot Market. Scientific Programming, 2022(1), 1540263.

[14] Liu, F., Cheng, G., & Pan, B. (2022). Smart management system in electric market considering customer demand response. Computers and Electrical Engineering, 99, 107830.

[15] Shang, C., Tian, Y., Yu, J., & Luo, Y. (2017, March). Application of intelligent power service system in power marketing. In 2017 7th International Conference on Education, Management, Computer and Society (EMCS 2017) (pp. 966-969). Atlantis Press.

[16] Zhao, C., Cheng, M., Chen, W., Wu, J., Xie, J., & Li, J. (2023, April). Personalized design of power marketing intelligent customer service system based on knowledge map. In Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022) (Vol. 12610, pp. 1059-1064). SPIE.

[17] Wu, S., & Dong, Z. (2022). An Auxiliary Decision-Making System for Electric Power Intelligent Customer Service Based on Hadoop. Scientific Programming, 2022(1), 5165718.

[18] Nie, J., Wang, Q., & Xiong, J. (2021). Research on intelligent service of customer service system. Cognitive Computation and Systems, 3(3), 197-205.

[19] Tan, Y., Xu, H., Wu, Y., Zhang, Z., An, Y., Xiong, Y., & Wang, F. (2021). Research on knowledge driven intelligent question answering system for electric power customer service. Procedia Computer Science, 187, 347-352.

[20] Gang Wu,Liang Hu,Yuxiao Hu,Yongheng Xing & Feng Wang. (2025). User intention prediction for trigger-action programming rule using multi-view representation learning. Expert Systems With Applications126198-126198.

[21] Alotaibi Mohammed & Hidayat-ur-Rehman Imdadullah. (2025). An empirical analysis of user intention to use chatbots for airline tickets consultation. Journal of Science and Technology Policy Management(1),204-228.

[22] Zhifeng Wei,Hongyan Wang,Qiang Xu,Yi Qu & Wei Xing. (2024). Optimized Attention Enhanced Temporal Graph Convolutional Network Espoused Research of Intelligent Customer Service System based on Natural Language Processing Technology. Applied Artificial Intelligence(1).

[23] Wu Zhong,She Qiping & Zhou Chuan. (2024). Intelligent Customer Service System Optimization Based on Artificial Intelligence. Journal of Organizational and End User Computing (JOEUC)(1),1-27.