



Hybrid LSTM-Transformer Modeling to Reconstruct Business Ecosystem Evolution Curves

Na Han^{1,*}

¹ School of Economics and Management, Zaozhuang University, Zaozhuang, Shandong, 277160, China

SUMMARY: *With the globalization of economy and the rapid development of technology, the business ecosystem has become an important support for the improvement of enterprise competitiveness and industrial development. In this paper, we constructed an LSTM-Transformer model combining Transformer, LSTM and multi-attention mechanism to simulate the evolution of business ecosystems, and carried out simulation experiments based on the takeaway O2O business model. The results show that the LSTM-Transformer model can fit the takeout platform data better, and its MAE, MAPE, and SMAPE values are smaller than those of CV-KF, ISTM, ISTM-Attention, and Transformer models, with the highest prediction accuracy and interpretability. Meanwhile, the simulation results show that advertising and promotion, word-of-mouth communication and the number of Internet users can affect the number of takeout platform users and their growth rate, product quality, logistics quality, the number of merchants, meal types, consumption level, and takeout pricing can affect the scale of the takeout market, and the trend of the evolution of the logistics capacity mainly depends on the demand for logistics capacity in the takeout industry. The study of the operation mechanism in the takeaway O2O business ecosystem in this paper helps to realize the healthy development of the takeaway O2O model business ecosystem and provides a reference for the operation of other business ecosystems.*

KEYWORDS: *transformer; LSTM; multi-attention mechanism; takeaway O2O model; business ecosystem*

1 Introduction

The rapid development of economic globalization and the increasing diversity of consumer demands have made it increasingly difficult for enterprises to survive and develop on their own. As in the biosphere, becoming the king of the forest and forming a herd may be a more correct choice than eating these small animals and becoming a bloated behemoth. Therefore, the alliance-related enterprises to create a business ecosystem suitable for their own development to take the road of steady win-win, which has become the mainstream trend of the current enterprise development [1, 2]. The concept of ecosystem was first proposed by the British ecologist A. G. Tansley, Wall Oddson on the basis of this definition of “organized behavioral systems”, that is, organized behavioral systems are ecological concepts in marketing performance, thus creating a precedent for the application of ecological ideas in the economic field [3].

Since the concept of business ecology was proposed, along with the deepening of enterprise

*15906375306@126.com

<https://doi.org/10.65102/is20261084>

diversification and the continuous exploration of enterprise business ecosystems, more and more scholars have carried out richer research around enterprise business ecosystems [4, 5]. For the way of ecosystem construction, Helfat and Raubitschek believe that business ecosystem is a dynamic network composed of innovation capability, environmental scanning and sensing capability, and coordination capability, which not only supports value creation, but also the way to support leaders to capture value [6]. In their study of business ecosystems, Lee, Y et al. pointed out that the main body, in addition to the participation of enterprises, consumers, and producers, also incorporates external stakeholders, government, and social organizations, which continuously enrich the members and constitute a set of complex network structure [7]. Peltoniemi research considers business ecosystem as a dynamic community destiny of interdependence, interaction and competitive cooperation of multiple members with the purpose of continuous innovation and success of business operations, which is influenced by the market environment [8]. Peltola, T sees the business ecosystem as a whole in which the value chains of individual firms, independent and intertwined, are a relatively stable structure [9]. Aarikka Stenroos and Ritala suggested that networks consist of participants that generate direct and indirect connections, and that business ecosystems include not only participants that generate direct and indirect connections, but also participants, technologies, and institutions that are interdependent in an informal and loosely organized manner [10].

In addition to researching the concept of business ecosystems, scholars have also conducted in-depth discussions and studies on the internal structure of business ecosystems and their characteristics. From the perspective of capabilities, Cennamo and Santaló argue that business ecosystems are superior to traditional vertically integrated systems because they have greater “generative capacity”, and that business ecosystems are more likely to foster complementary innovation capabilities from autonomous, heterogeneous firms [11]. Burford, S and Resmini, A conducted an in-depth study on the causes and purposes of business ecosystems and found that in digital business environments, firms can make more profit, in less time, at lower cost, and improve the stability and resilience of business ecosystems [12]. Mäkinen, S J et al. suggested that the competitiveness of a platform-based business ecosystem comes from the improvement of the services of the individual members, and that cooperation, competition, and co-evolution among firms in this system can improve the competitiveness of the firms [13]. Fujita, M et al. pointed out in their study that in business ecosystems, there are characteristics of diversification and decentralization, so constructing a business ecosystem can reduce the operational risk of the enterprise to a certain extent and improve the enterprise's ability to cope with external risks [14]. Cui, Z et al. further analyzed the role of business ecosystems in enterprises, pointing out that the application of business ecosystems in enterprises can effectively improve the operational efficiency of enterprises, increase their competitiveness and profit margins, and at the same time strengthen the internal business environment control ability and external business environment response ability of enterprises [15].

The establishment and application of business ecosystems is ultimately aimed at serving enterprises, i.e. realizing enterprise value creation, so scholars have further analyzed business ecosystems from the perspective of business ecosystems and enterprise value creation [16-18]. From the perspective of business ecosystem to realize the effect of enterprise value creation, Pera, R et al. believe that value co-creation within the business ecosystem is a multi-party interactive and dynamic development process, and in the whole business ecosystem, the interactive subject of value co-creation is not only the user and the enterprise, but also should include at least two economic activities participating subjects [19]. Rong, K et al. argue that business ecosystems form where groups of specialized firms around platform-based firms, through network effects and co-evolution in order to jointly cope with increasingly complex external competitive environments, and that the key to their success lies in the co-evolution of

stakeholders in the ecosystems and the co-creating of value with customers [20]. Nuutinen, M et al. proposed value co-creation as a dynamic synergistic mechanism of service interaction and resource integration, which is strategically important in the theoretical framework of business ecosystems, and the paradigm breaks through the traditional enterprise boundaries and realizes a fundamental shift in the logic of value generation from linear creation to network symbiosis through cross-principal resource orchestration and demand synergy [21]. Kumar, P et al in their study stated that business ecosystems based on value co-creation should focus on the orientation of the system, application of organizational theories, organizational innovation, networking capabilities of the system structure and processes, improvement of the value sharing mechanism, co-created value and measuring and value [22]. Adner, R, on the other hand, argues that business ecosystems are dynamic systems composed of multilateral partners, including stakeholders such as upstream suppliers, downstream customers, shareholders, trading partners, and governments, through communication and interaction, with value co-creation as a core feature [23].

From the perspective of business ecosystem to realize the path of enterprise value creation, Clarysse, B et al. pointed out that different enterprises which are competitors to each other under the traditional business model can also achieve win-win cooperation in business ecosystem, so in the process of business ecosystem operation, in order to make the enterprise better development, it is necessary for more enterprises to participate in it to strengthen the collaboration, and then optimize the business ecosystem [24]. Awano, H and Tsujimoto, M analyze the value creation mode of business ecosystem and propose that in this mode, due to the characteristics of complementarity among the members of the business ecosystem, the enterprise can use the business ecosystem to expand the advantages of technological leadership, production scale, and integrity of the operation mode, which can be realized through the transfer of patents and sharing within the system to achieve the enhancement of the value creation capacity of the business. Through patent transfer and sharing within the system, the value creation ability of the members of the ecosystem can be enhanced, thus realizing the enhancement of the value creation ability of the enterprise as a whole [25]. Alet, J stated that a key feature of business ecosystems is value co-creation, i.e., the co-creation of value through collaboration between member firms and between firms and their external environment for mutual benefit [26]. Felch, V and Sucky, E believe that a business ecosystem can be regarded as a network composed of multiple participants, who jointly create value through collaboration or competition. The core lies in the interdependence among members and the goal of co-evolution. And it is believed that "co-created value propositions" or "shared destiny" are key characteristics of business ecosystems [27].

The evolution of business ecosystems is a complex and dynamic process, and the path of business ecosystems has been summarized by academics as an initial pioneering period, a subsequent expansion period, a leadership formation period, a self-reinvention and a decline period [28, 29]. For the evolutionary path of business ecosystems, Jacobides, M G et al. argued that modularity provides the conditions for the emergence of business ecosystems, so that core firms tend to integrate more partners as the life cycle evolves in order to reduce the uncertainty of partner behavior [30]. Bertassini, A C et al. pointed out that the strategic choices of enterprise subjects in business ecosystems at different stages of development are characterized by environmental adaptability, and their decision-making mechanism dynamically adjusts the resource allocation mode and evolutionary path through sensing changes in external ecological conditions and the reconstruction of competitive relationships, in order to maintain the sustained kinetic energy of their organizational development [31]. Yoon, C et al. pointed out that core firms will bear the costs and risks of developing new business ecosystems in the early evolutionary stages of business ecosystems and have low initial bargaining power, but as

business ecosystems evolve over time, core firms will receive significant investments to develop business ecosystems and may be able to adapt to newly created business ecosystems by adapting their existing business models [32]. Kapoor and Lee argue that forms of business alliances provide greater incentives than hierarchies, and that the difference between the evolutionary paths of business alliances and business eco-networks is that alliances are concerned with how firms can benefit from dichotomous relationships, while eco-networks are more concerned with how firms can benefit more from the structure of the dichotomous relationships that are forming a network with multiple firms [33].

In this paper, a fusion model based on LSTM and Transformer is constructed. Firstly, the short-term features of the data are extracted using LSTM, then the output of LSTM is used to further extract the long-term features using Transformer, and finally the final prediction results are generated through the fully connected layer. In order to test the effectiveness of the model, experimental comparisons of its predicted and actual values are made, then performance comparisons are made with other models, followed by an analysis of the interpretability of the model, and finally a simulation analysis is conducted to deeply explore the evolutionary trends of the number of platform users, the number of merchants, the market size, the logistics capacity and their key factors in the takeaway O2O business ecosystem.

2 Hybrid LSTM-Transformer model construction

In this paper, the hybrid LSTM-Transformer model is used to reconstruct the business ecosystem evolution curve, and the construction process of the LSTM-Transformer model is specified in this chapter.

2.1 Overall structure of the LSTM-Transformer model

The overall architecture of the LSTM-Transformer model proposed in this paper is shown in Fig. 1, which consists of four main parts: the LSTM layer, the location coding layer, the self-attention layer, and the fully connected layer. Its topmost input is a commercial ecosystem time series, and the length of the sequence can be self-defined to represent the number of time steps in the input, i.e., the number of historical data $m - 1$. Each time step in the sequence is a vector with a vector length of 5, which consists of the number of users, market size, number of merchants, logistics capacity, and number of orders of the business ecosystem, representing the comprehensive information of the business ecosystem at each point in time. In order to improve the problem of Transformer's weak ability for short-term temporal feature extraction and the loss of temporal order in parallel computing mode, LSTM is utilized for the initial extraction of temporal features. Then the output features of LSTM are positionally encoded using sine and cosine functions, and the encoded features are input into the self-attention layer for parallel feature extraction. Finally, the features extracted from the LSTM and the self-attention layer are linearly transformed by the fully connected layer, and the final output is the predicted sequence m .

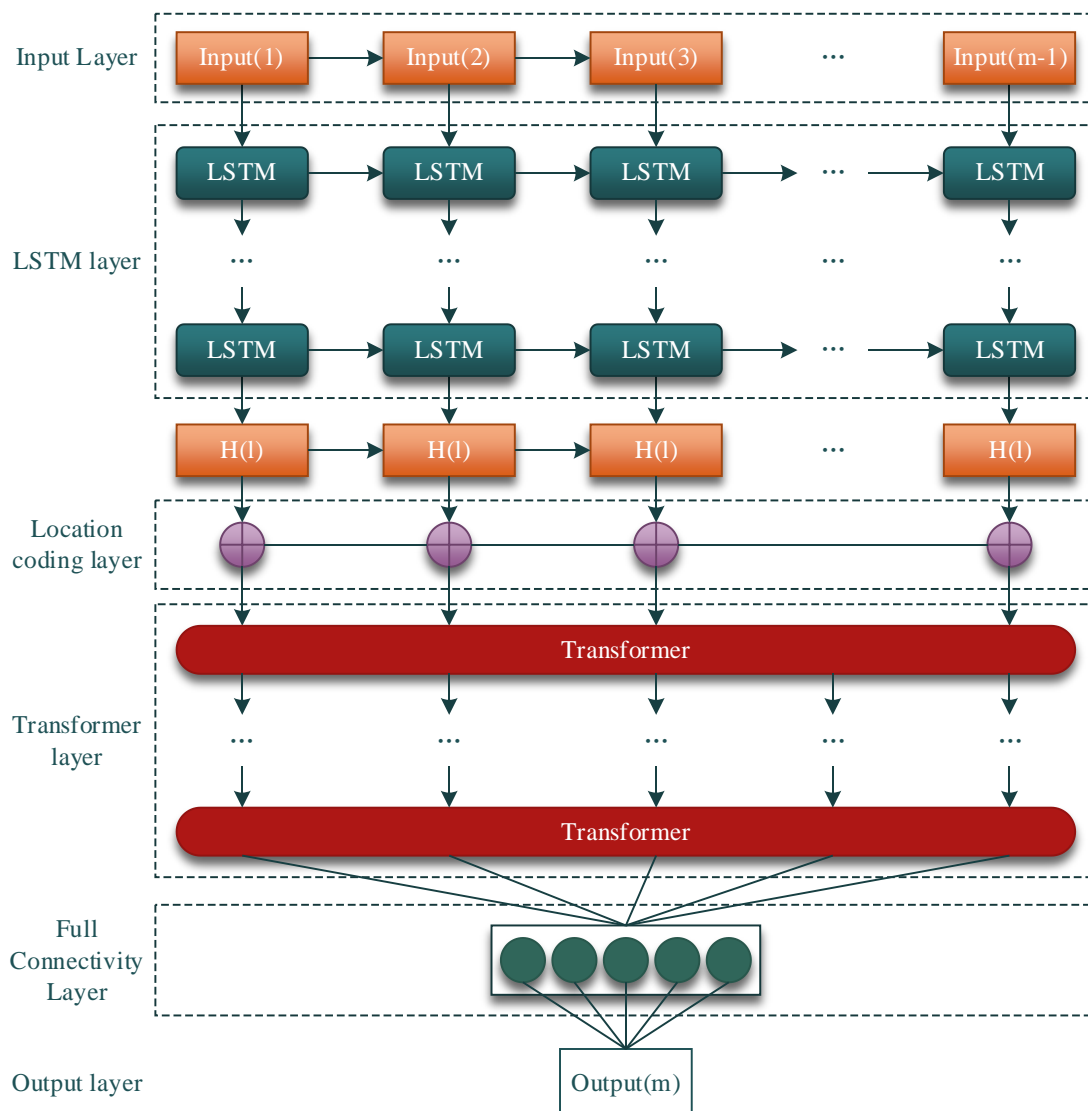


Figure 1: Architecture of LSTM-Transformer model

2.2 LSTM layer

To address the problem of gradient vanishing or gradient explosion in recurrent neural networks (RNNs), related scholars have proposed the Long Short-Term Memory Neural Network (LSTM). The LSTM is an optimized variant of the RNN, and the structure of this network, due to the introduction of a cellular state used to carry out linear feedback transfer, can learn long term dependency information, has the ability to filter historical and existing information, and is more conducive to time series data Processing.

The key to LSTM is the cell state. Cell states are similar to conveyor belts in that information can be passed and flow along the complete chain during cell state operation except for a small number of linear interactions. LSTM does have the ability to remove or add information to cell states, regulated by structures called gates, a gating mechanism that can be effective in solving the problem of vanishing and exploding gradients in RNNs. LSTM has three such gates, oblivion gates, input gate and output gate to protect and control the cell state.

(1) Oblivion gate

The first is the “forgetting gate”, the structure that decides to forget part of the information from the current cell state. The oblivion gate reads the previous output h_{t-1} and the current

input x_t , makes a sigmoid nonlinear mapping, and outputs a vector f_t , which is multiplied by the cell state C_{t-1} . The output vector is mapped between 0 and 1 by the sigmoid function, where 0 means completely discarded and 1 means completely retained. The output vector f_t of the forgetting gate is denoted as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where W_f is the weight matrix of the forgetting gate and b_f is the bias vector.

(2) Input gate

The next step is to determine the new information to be stored in the cell state. Storing new information is divided into two parts: the first is the “input gate”, a sigmoid layer in the structure that determines which values will be updated. The second is the tanh layer, which creates the vector C_t of new candidate values and adds them to the state. The formula is shown below:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

(3) Cell state renewal

Update the cell state, in which C_{t-1} will be updated to C_t . The old cell state C_{t-1} is multiplied by f_t to determine the information to be discarded, plus $i_t \times \tilde{C}_t$. C_t i.e., the new cell state candidate value, which changes according to the degree of state update. The cell state update process is shown below:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

(4) Output Gate

The final output value of the “output gate” will be output based on the cell state and certain weights. First, a sigmoid layer is run to determine the output portion of the cell state. Then the cell state C_t processed by the tanh layer is multiplied with the output of the sigmoid layer, and the final output is h_t :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

Up to this point, the LSTM cell at a given moment has gone through a series of workflows of forgetting gates, input gates, cell updating states, and output gates to complete the process, and the LSTM model will repeat this process during training until the end of training.

2.3 Transformer layer

The structure of the Transformer is shown in Fig. 2. The overall structure of the Transformer follows the Encoder-Decoder framework, where both the Encoder and Decoder layers consist of exactly the same N layers containing Attention. Each Attention layer in the Encoder layer

contains two sub-layers, namely Multi-head Attention mechanism and Fully Connected Feed-forward Neural Network, and the residuals of the data are connected and normalized at each step. The Decoder layer, on the other hand, in addition to the two sub-layers mentioned above, has another layer of Multi-head Attention for handling the processing of the outputs from the Encoder layer for Mask masking the output from the Encoder layer, which is similar to the Decoder layer, also performing residual concatenation and normalization after each sublayer.

Transformer's encoder has a better ability to learn time-series representations and can extract time-series features, so the design of this paper only uses the encoder part of Transformer as a time-series feature extractor to extract time-series features of business ecosystems.

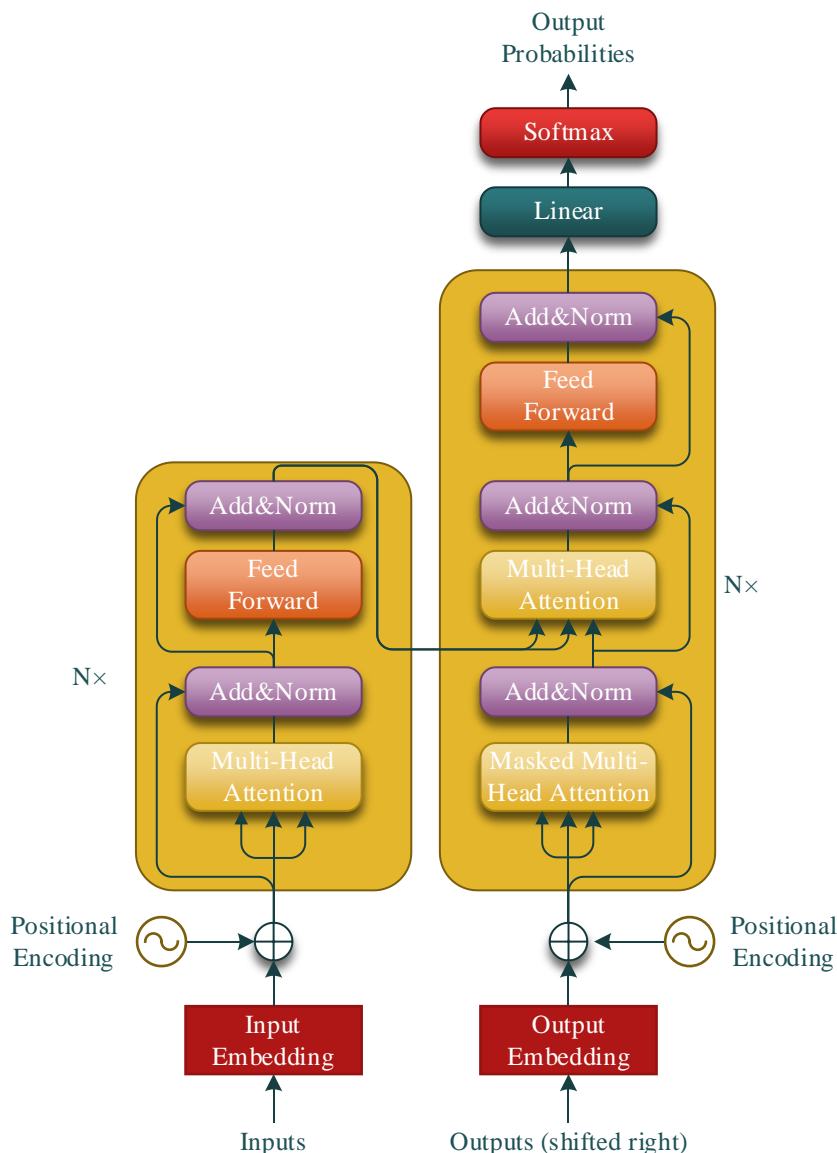


Figure 2: Transformer structure

2.3.1 Location coding mechanisms

Since the structure of Transformer model does not contain recurrent neural network and convolutional neural network, in order to make the model able to deal with the sequential information in the sequence, the Transformer model adds positional encoding in the input

embedding layer of Encoder and Decoder to mark the relative position of the input data, the mathematical formula for positional encoding is as follows:

$$PE_{pos,2i} = \sin \frac{pos}{10000^{\frac{2i}{d_{model}}}} \quad (7)$$

$$PE_{pos,2i+1} = \cos \frac{pos}{10000^{\frac{2i}{d_{model}}}} \quad (8)$$

where pos denotes the position where the word vector is located. i denotes the word vector dimension. d_{model} denotes the dimension of the position vector, which is equivalent to the dimension of the model. PE is of the form of a matrix, and $PE_{pos,2i}$ denotes the value of the pos th row, $2i$ th column of the matrix. $PE_{pos,2i+1}$ then denotes the value of the pos row, $2i+1$ column.

2.3.2 Attention mechanisms

The attention mechanism is utilized to calculate the weights of the attention distribution of the input variables. The new input variables are then determined using the resulting attention distribution.

The input of Attention consists of the vectors query, key and value, by applying the Attention function the vectors are transformed into matrix Q , matrix K and matrix V , as in equation (9) the output matrix is calculated:

$$Attention(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_k}} V \right) \quad (9)$$

2.3.3 Multi-pronged self-attention mechanisms

Compared to the single execution of an Attention mechanism, it was found to be more beneficial to perform the process of converting the input vector to a matrix h times using different linear mappings, where the output values of the parallel Attention mechanisms are spliced together and mapped again to obtain the final value. The above process is the Multi-head Attention mechanism, which is mathematically expressed in equation (10):

$$\begin{aligned} MultiHead(Q, K, V) &= concate(head_1, \dots, head_h)W^o \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (10)$$

where $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_k}$.

2.3.4 Residual connections

As the depth of the model continues to deepen, its backpropagation when the gradient transfer is more and more difficult, which can lead to gradient effect or explosion problem, so the residual network (ResNet) is proposed to go to solve the gradient problem in deep neural networks.

The basic structure of a residual network is shown in Figure 3. The residual network is based on the hopping mechanism. x skips certain layers through some paths, which are called “hopping connections”. By jumping connections, it is possible to make changes in what is learned in certain layers.

Without the residual structure, the input of this network is x and the output is $F(x)$. The network learns the weights W_F and b_F , at which point $F(x)$ is shown below:

$$F(x) = Relu(W_F \cdot x + b_F) \tag{11}$$

After adding the residual structure, the network becomes:

$$H(x) = F(x) + x \tag{12}$$

Jump connections in ResNet solve the problem of gradient degradation in deep neural networks due to too many layers by allowing the gradient to flow through the path established by the jump connections, propagating the gradient back to the lower layers. Another advantage of jump connections is that they allow the model to learn constant functions, ensuring that the upper layers are at least better characterized compared to the lower layers, not worse.

To deal with the problem of vanishing gradients in neural networks more effectively, the model is trained more efficiently. In the Transformer layer, a residual connection structure is used between the self-attentive module and the fully connected module with the formula:

$$X = X_{input} + attention(Q, K, V) \tag{13}$$

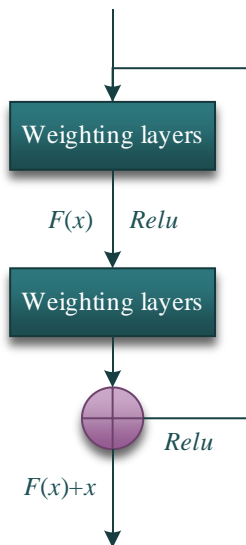


Figure 3: Residual module

2.3.5 Standardization

In the Transformer layer, normalization is used simultaneously between the self-attentive module and the fully connected module to accelerate training. The main function of normalization, which is to regress all the hidden variables of the model to a standard normal distribution so that the model can accelerate convergence during training, is calculated as follows:

$$\mu_j = \frac{1}{m} \sum_i^m x_{ij} \quad (14)$$

$$\sigma_j^2 = \frac{1}{m} \sum_i^m (x_{ij} - \mu_j)^2 \quad (15)$$

$$\text{LayerNorm}(x) = \frac{x_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \varepsilon}} \quad (16)$$

where x_{ij} is the value of the i th row and j th column of the output matrix, μ_j is the mean of the j th column of the output matrix, and σ_j^2 is the variance of the j th column of the output matrix. The standardized result is obtained by subtracting all the elements of each column from the mean of that column and dividing by the standard deviation of that column. To avoid the denominators all being zero, a constant ε that is not zero is added to the denominator part.

2.4 Full connectivity layer

The fully connected layer in the model of this paper utilizes a BP neural network. The structure of the fully connected layer is shown in Fig. 4, which consists of three main parts, i.e., input layer, hidden layer and output layer. Among them, there is a many-to-many relationship between layers. The BP neural network learns complex nonlinear relationships through neuron superposition and nonlinear activation functions. When the output data is input to the input layer of the network, and the input layer data is input to the hidden layer, the data will be activated by weight calculation and nonlinear activation function, and then the hidden layer data is input to the output layer to get the final output data, and the data in the hidden layer to the output layer is also subject to weight calculation and nonlinear activation.

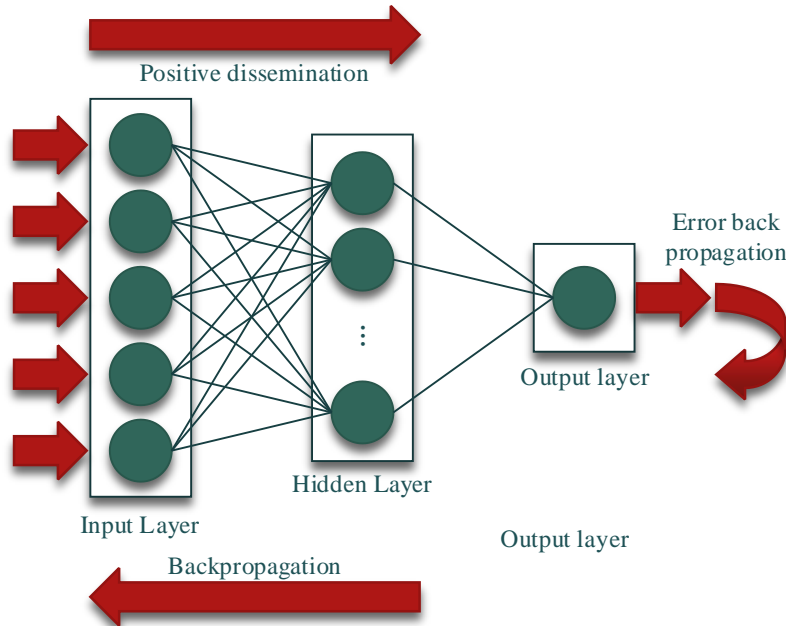


Figure 4: Fully connected layer structure

The neurons in the BP neural network have two hyperparameters, weight and deviation, and can calculate the error by comparing the output value obtained from the forward calculation with the actual value, and then back-propagate according to the gradient, and make the error of the whole network smaller by changing the gradient of the neuron with larger error gradient. Finally learning to get a complex mapping between input and output.

2.5 Model training

2.5.1 Optimizer

In this paper, adaptive moment estimation Adam is chosen as the optimizer, which can dynamically adjust the learning rate of each parameter since it is itself on the RMSprop with a momentum term, using first-order moment estimation with second-order moment estimation of the gradient. The advantage of Adam mainly lies in the fact that after calibrating it by bias, there is an update of the number of times of learning in each iteration, which makes the parameter smoother.

Where the first order momentum is calculated as:

$$m_t = \alpha m_{t-1} + (1 - \alpha) \nabla C(w) \quad (17)$$

The second order momentum is given by:

$$v_t = \beta v_{t-1} + (1 - \beta) \nabla C^2(w) \quad (18)$$

The parameter iterative update formula is:

$$w_{t+1} = w_t - \eta \frac{\nabla C(w)}{\sqrt{s_t + \epsilon}} \quad (19)$$

where α usually takes the value of 0.9, β usually takes the value of 0.999, and ϵ usually takes 10^{-8} . Since the Adam method combines the respective advantages of the SGD method and the RMSprop method, in practice, Adam is the most commonly used method to obtain a prediction result relatively quickly.

2.5.2 Activation functions

A better activation function improves the nonlinear capability of the model and also affects the backpropagation to a lesser extent.

The common activation functions are Softmax, Relu, Tanh, Sigmoid, etc. In this paper, Relu function is used as the activation function. In this paper, Relu function is used as the activation function. Relu is actually the function that sets 0 as the threshold to take the maximum value, and its mathematical expression is shown below:

$$Relu = \max(0, x) \quad (20)$$

2.5.3 Loss Functions

The loss function is an arithmetic function used to measure the extent to which the predicted value $f(x)$ of a model differs from the actual value Y . The loss function is usually a non-negative function, and the smaller the value of the function, the closer the model's results are to

the actual values.

The loss function is applied during the training phase of modeling, when each batch of input values is fed into the model, first through forward propagation, the model derives the corresponding predictions, and then the loss function is used to derive the difference values. Finally, the back-propagation mechanism is then used to update these discrepancy values to the parameters in the model, thus enabling the model to make more accurate predictions.

For time-series prediction tasks, the mean square error (MSE) is often used as a regression loss function to avoid the influence of outliers on the model, thus improving the training efficiency of the model. The MSE is calculated as the sum of the squares of the differences between the expected and actual numbers.

3 LSTM-Transformer based business ecosystem evolution

In this chapter, taking the takeaway O2O business ecosystem as an example, we first use the LSTM-Transformer model to predict the total customer demand, and at the same time compare other models to test the model validity and carry out the interpretability analysis, and finally carry out the simulation analysis of the evolution trend of the takeaway O2O business ecosystem.

3.1 Data set acquisition

The data used in this paper mainly comes from all the real order data of a takeaway platform occurring in city D during the period from May 1, 2020 to April 30, 2025. The platform received a total of 3,964,826 takeout order creation information in this 5-year period, with a total of 72 fields. The fields mainly include user-related information such as user's name, gender, delivery address, merchant-related information such as merchant's name, business district, order creation time, order status, order amount and other order-related information.

3.2 Model Training Setup and Testing Process

In this paper, the input data of the model is a vector of length 5, the number of hidden layers of coding layer is set to 512, the number of hidden layers of LSTM layer is set to 128, 1 layer each of coding layer and LSTM layer, the number of attention head is set to 4, the learning rate is 0.0001, the batch processing is used, the batch size of the network training is 64, and the ratio of the training set to the test set is 7:3, the network training is done by using the Adam optimization algorithm, the loss function uses the mean square error loss (MSEloss), training 800 rounds.

The LSTM-Transformer model training error variation curve is shown in Figure 5. With the increase in the number of training times, the model error gradually decreases, and finally convergence is achieved, indicating that the model training effect is good.

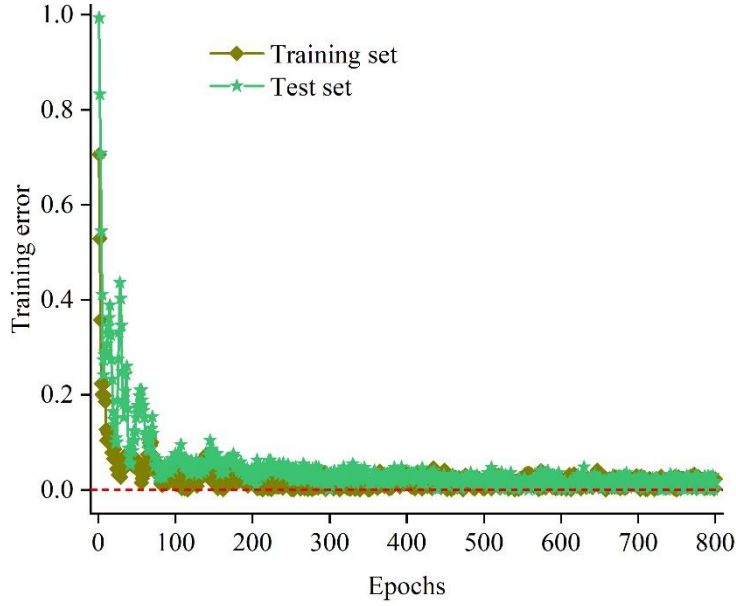


Figure 5: Model error variation curve

3.3 Evaluation indicators

In this paper, three commonly used metrics are used to assess the error of the prediction model, including mean absolute error (MAE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE), with the following formulas:

$$MAE = \frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i| \quad (21)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{h(x_i) - y_i}{y_i} \right| \quad (22)$$

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{h(x_i) - y_i}{\left[\frac{h(x_i) + y_i}{2} \right]} \right| \quad (23)$$

where y_i denotes the true value of the predicted data, $h(x_i)$ denotes the predicted value, and n denotes the number of samples.

3.4 Validation of model validity

3.4.1 Experimental results and analysis

In order to illustrate the prediction effect of the proposed LSTM-Transformer model, the total customer demand observations of the sample takeout platform in May 2024 are plotted as shown in Figure 6. It can be seen that the deviation between the predicted and actual values of the LSTM-Transformer model is very small, indicating that the model can predict the total customer consumption demand of the takeaway platform well, i.e., the proposed model is effective.

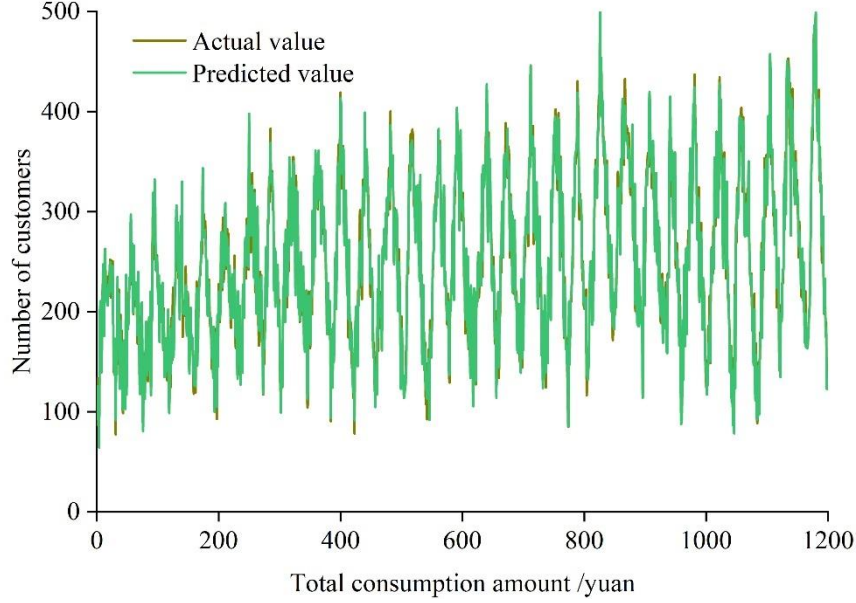


Figure 6: The total customer demand observation results of the food delivery platform

3.4.2 Model Performance Comparison

In this paper, the traditional constant velocity Kalman filter model (CVKF), the traditional LSTM model, the LSTM-Attention model with the Attention mechanism added and the Transformer coding layer model are selected for comparison to verify the high accuracy and strong robustness of the model in this paper.

Comparison of the performance of different models with single-step prediction step length is shown in Table 1, the initial step length $t_0=5$ days is selected, and this paper analyzes the prediction accuracy of the models by comparing the error of different models in predicting the consumption amount of customers in the next time period.

Compared with CV-KF, LSTM, LSTM-Attention and Transformer, LSTM-Transformer has the smallest MAE, MAPE and SMAPE values, which are 4.124 yuan, 20.316% and 22.057% respectively. Thus, it shows that the LSTM-Transformer model has the highest prediction accuracy and outperforms CV-KF, LSTM, LSTM-Attention and Transformer in single-step prediction.

Table 1: Comparison of single-step prediction step size performance of different models

Model	MAE /yuan	MAPE /%	SMAPE /%
CV-KF	9.257	52.532	33.941
LSTM	6.038	32.418	34.536
LSTM-Attention	5.893	30.624	32.843
Transformer	6.571	39.283	39.655
LSTM-Transformer	4.124	20.136	22.057

Predicting multi-step length is beneficial for planning marketing strategies in advance so that the platform can boost the number of orders and transaction amount. The validation of the model for multi-step prediction can test the stability of the model in terms of long time series prediction. The prediction errors in the case of different methods with different step lengths are shown in Table 2.

It can be seen that with the increase of the step length of predicting the consumption amount, the LSTM model, the LSTM-Attention model and the Transformer model have a larger growth

in error, while the LSTM-Transformer model utilizes the LSTM to strengthen the training after the Transformer module has deeply mined the trajectory features of the different levels, which makes the prediction error keep a more The LSTM-Transformer model is trained with LSTM after deeply mining different levels of trajectory features in the Transformer module, which keeps the prediction error smaller, has stronger model stability and higher prediction accuracy, and has obvious advantages in the prediction of long sequence trajectories.

Table 2: Prediction errors under different methods and different synchronization lengths

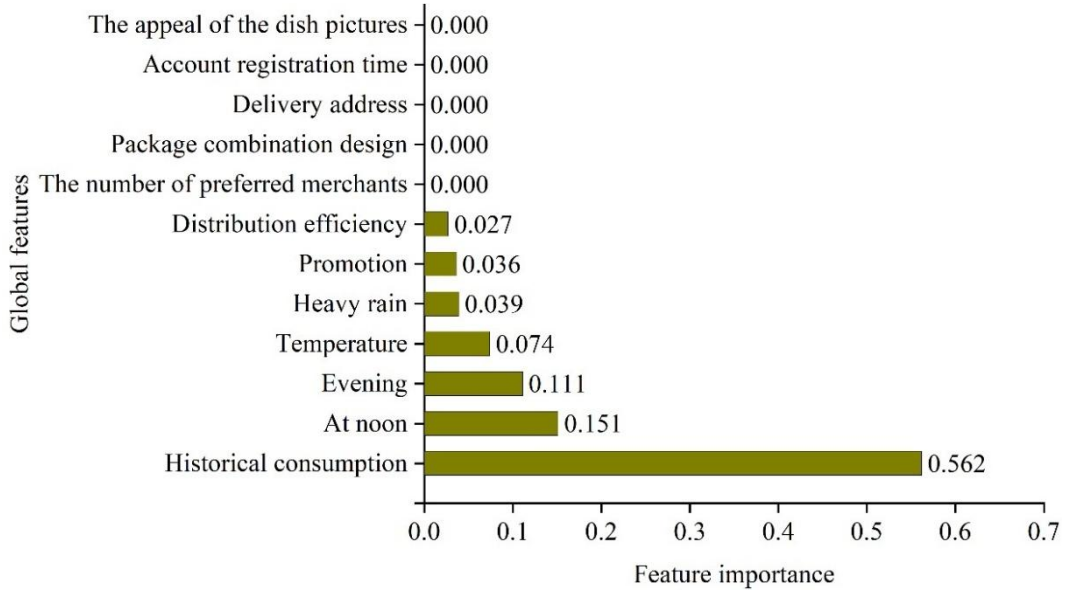
Error index	Model	Prediction step size			
		2	3	4	5
MAE /yuan	LSTM	9.127	13.459	15.873	18.324
	LSTM-Attention	8.415	12.247	14.378	17.973
	Transformer	8.372	10.041	11.658	14.221
	LSTM-Transformer	5.326	5.508	6.291	6.632
MAPE /%	LSTM	34.039	35.741	37.528	39.404
	LSTM-Attention	32.155	33.763	35.451	37.224
	Transformer	41.247	43.310	45.475	47.749
	LSTM-Transformer	21.143	22.200	23.310	24.475
SMAPE /%	LSTM	36.263	38.076	39.981	41.979
	LSTM-Attention	34.485	36.209	38.022	39.923
	Transformer	41.638	43.720	45.906	48.201
	LSTM-Transformer	34.039	35.741	36.928	38.404

3.4.3 Interpretability analysis

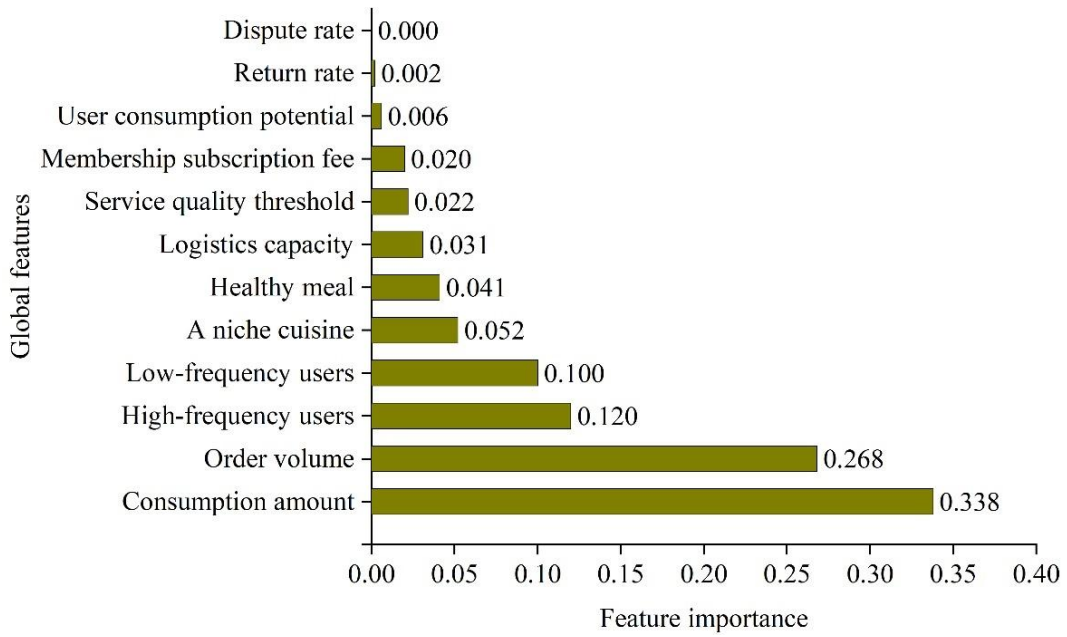
In this section, we analyze how LSTM-Transformer interprets the predictions. Firstly, we analyze the feature importance learned by LSTM-Transformer during the training process. The global feature importance computed by LSTM-Transformer is shown in Fig. 7, and (a)~(b) denote the conditional window and prediction window, respectively. The vertical axis represents the input variables and the horizontal axis is the global feature importance score.

In the conditional window, as expected, past observations of consumption are the most relevant with a share of 0.562. Historical observations of consumption are the most direct reference for forecasting, and therefore the most important input variable. In addition, time variables such as midday and evening and weather variables such as temperature and heavy rainfall also contribute to the model predictions, accounting for 0.262 and 0.113, respectively. In fact, they allow the model to make assumptions about current environmental conditions. In contrast, the other variables have little value in the conditional window.

In the prediction window, consumption behavior and user characteristics variables are known inputs. Among them, the contribution of the amount of consumption amounts to 0.338, which is the most relevant input, and it is directly related to the customer's consumption needs. Order amount is the second relevant variable with a contribution of 0.268 and it affects the total amount spent by the customer. Therefore, obtaining the accurate amount of consumption and order quantity is very important for consumption demand. In addition, high-frequency users is the third most important variable with a contribution of 0.120, which reflects the stable flow of the takeout platform. It can also be seen from Figure 7 that the model extracts information from only a subset of input variables to generate predictions, suppressing irrelevant variables so that they do not affect the model predictions.



(a) Condition window



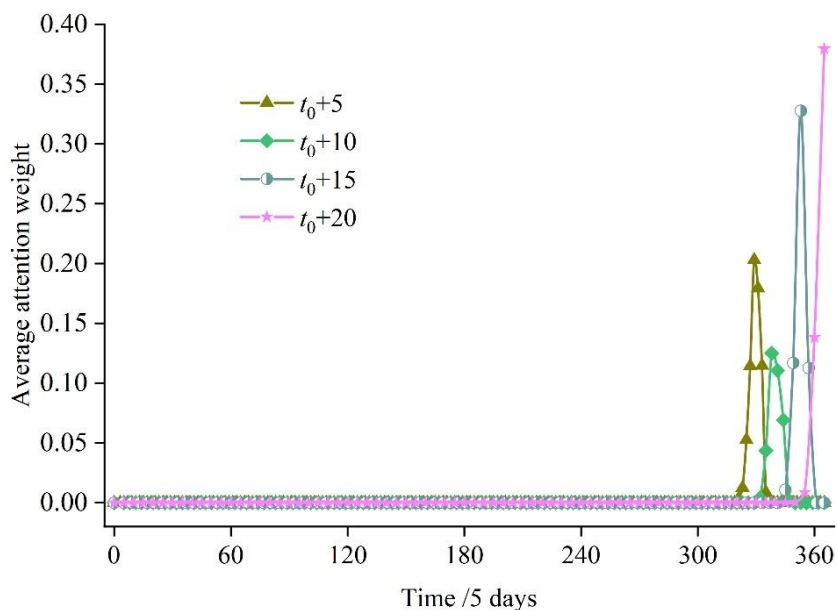
(b) Prediction window

Figure 7: Global feature importance learned by LSTM-Transformer

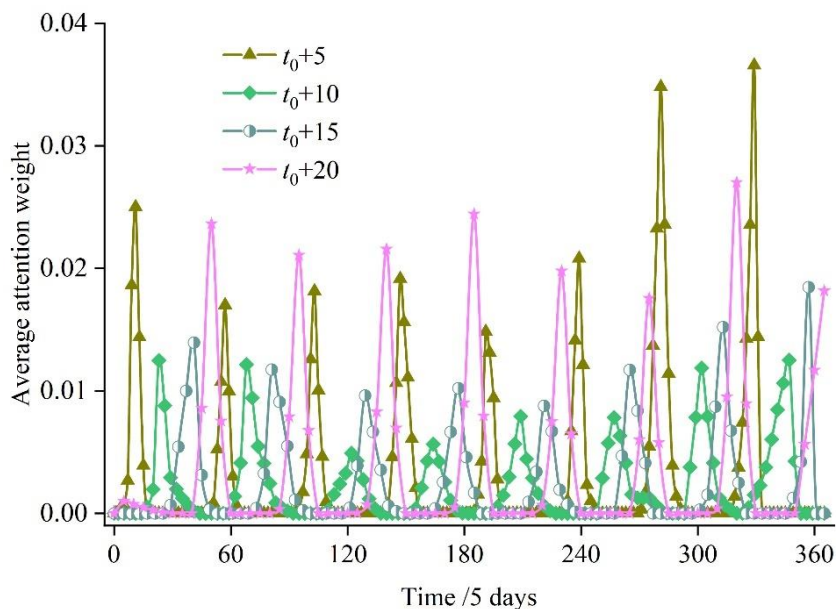
Attentional weight patterns can reveal the most important time steps for making predictions. LSTM-Transformer learns these temporal patterns from the original training data through an interpretable multi-head self-attention layer. The attention weight patterns of the first and last layers of the prediction model are shown in Fig. 8's (a)~(b). This section visualizes the average attention weights for different prediction time steps on the training set, i.e., $t \in \{t_0 + 1, t_0 + 2, t_0 + 3, t_0 + 4\}$.

In the first layer, it can be observed that the attention weights are only generated within the prediction window, and the closer to the prediction time step, the higher the weights. This means that the model first extracts information about the prediction time step of the current window.

In the last layer, the model attends to all inputs of the conditional and prediction windows. The attentional weights show a seasonal pattern, with peaks of attention occurring at essentially equal intervals on different days. The LSTM-Transformer learns these temporal dependencies directly from the input data, without requiring any artificial assumptions. This allows the developer to easily see where in the input sequence the model is focusing attention in each layer at each prediction time step. This capability can be very helpful in improving the credibility of the model. Alternatively, developers can use these insights to check whether the model captures the true causal relationship between inputs and outputs.



(a) The first floor



(b) The last layer

Figure 8: The time patterns learned by LSTM-Transformer

3.5 Simulation Analysis

3.5.1 Trends in system evolution

After the LSTM-Transformer model passed the validity test, the model was used to simulate the evolutionary trend of the main population indicators in the takeaway O2O business ecosystem, and the simulation time was set to May 2020-April 2025, in which the data from November 2020-April 2025 were the predicted values.

The evolutionary trend of the number of takeaway users is shown in Figure 9, which shows that the statistics and forecasts are made every 6 months, and the number of takeaway users shows an incremental trend, and its growth rate from May 2020-May 2022 has also been increasing, but from November 2022 onwards the growth rate of the number of takeaway users began to decline year by year. This is because after the initial explosive growth of the takeaway O2O industry, the growth rate of takeaway users began to slow down due to the limitations of the Internet penetration rate and the diminishing role of the demographic dividend. According to the change curve of the growth rate of takeaway users, the number of takeaway users will still continue to grow in recent years, but it may enter a relatively smooth evolutionary stage in the future.

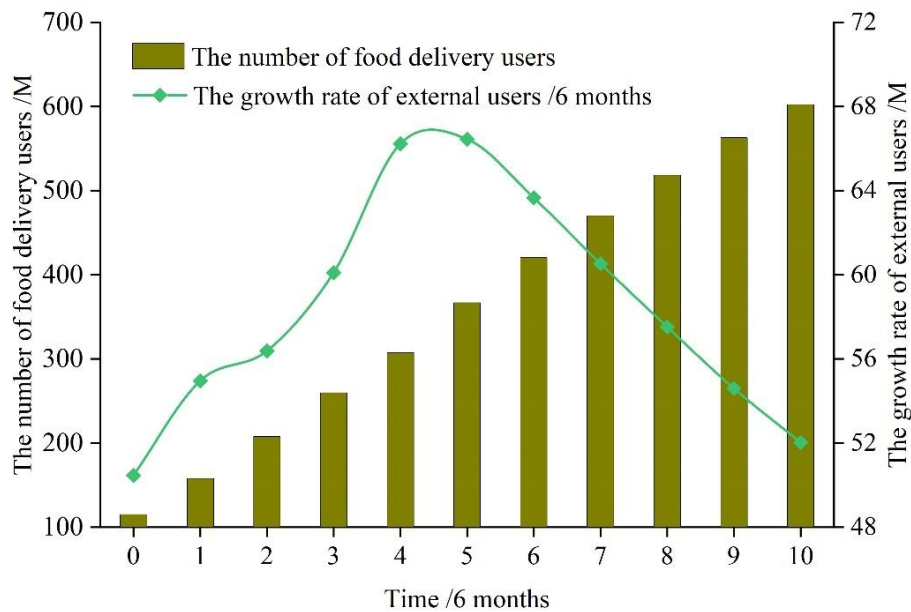


Figure 9: The evolution trend of the number of food delivery users

The evolutionary trend of the takeaway market size is shown in Figure 10, in which the takeaway market size also maintains a continuous growth trend, but its growth rate after November 2020 experienced a broken-bank decline. This is mainly due to the impact of the new crown epidemic in 2020, during which the consumption of residents and the transportation of materials were greatly affected, causing the catering industry to suffer a huge blow, and ultimately affecting the growth rate of the takeaway market size.

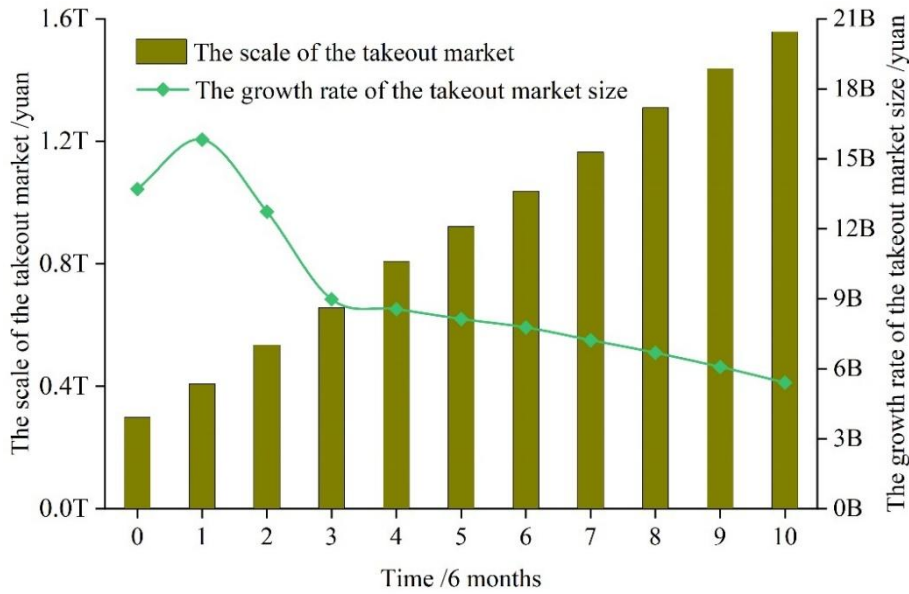


Figure 10: Evolution trend of the takeout market size

The trend of changes in the number of merchants is shown in Figure 11, the number of merchants has been in a state of growth during the period of 2020-2025, but the growth rate of the number of merchants appeared a more obvious decline after November 2020, which is mainly due to the fact that the role of the Internet traffic dividend is gradually becoming smaller, and with the arrival of the information age, the threshold of entry into the food and beverage industry is getting higher and higher, coupled with the impact of the new crown epidemic, the food and beverage With the impact of the new crown epidemic, the transportation of related materials and ingredients was blocked, the pressure of rent increased, and the financial chain of many catering merchants had problems and could not operate normally, resulting in a gradual slowdown in the growth of the number of merchants in the takeaway O2O industry.

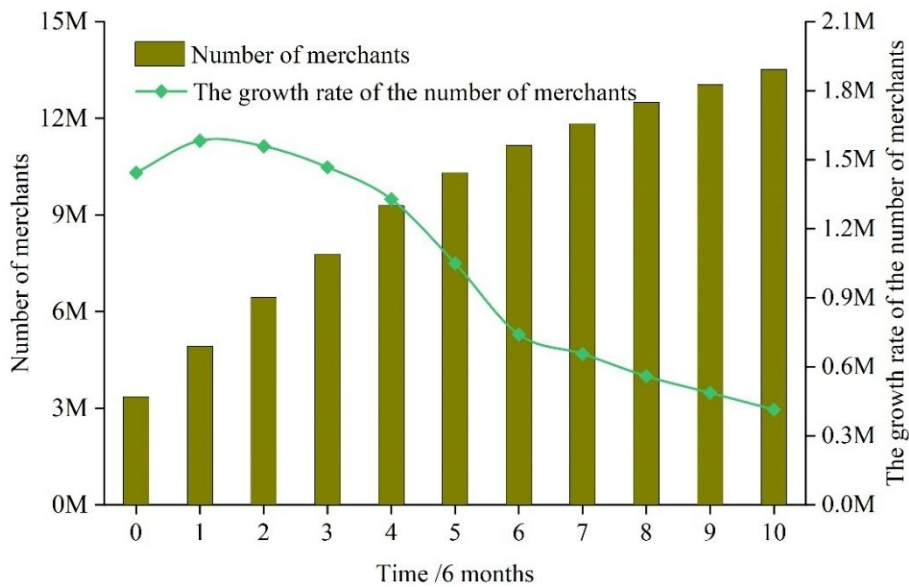


Figure 11: Trend of changes in the number of merchants

The trend of logistics capacity evolution is shown in Figure 12. It can be seen that there is a very obvious increase in the growth rate of logistics capacity between May 2021 and May

2023, which is due to the fact that the growth rate of the takeaway market size increased significantly from May to November 2020, but the logistics capacity increased significantly from 2021 due to a certain delay in the effectiveness of logistics construction. However, by around November 2023, the growth rate of logistics capacity shows a gradually decreasing trend due to the gradually decreasing supply-demand contradiction between the takeaway market size and logistics capacity.

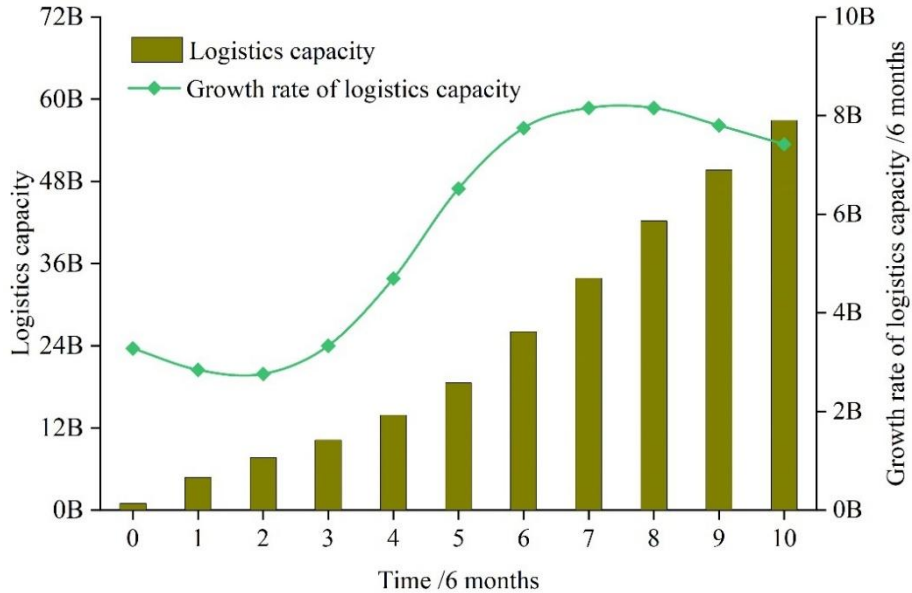


Figure 12: Evolution trend of logistics capacity

3.5.2 Simulation analysis of key factors

In order to further study the key factors affecting the evolution of the takeaway O2O business ecosystem, this paper takes the number of takeaway users as an example, simulates its key factors, and observes the degree of influence of its changes on the evolution of population indicators within the takeaway O2O business ecosystem by changing the magnitude of the values of the key factors, i.e., analyzes the sensitivity of the level variables to the relevant constants.

In the takeaway O2O business ecosystem, changes in advertising promotion, word-of-mouth communication and the number of Internet users will have an impact on the evolution of the number of takeaway users. In this paper, we make the following adjustments to the parameters: the advertising promotion factor is adjusted upward by 20%, the advertising promotion factor is adjusted downward by 20%, the word-of-mouth communication factor is adjusted upward by 20%, the word-of-mouth communication factor is adjusted downward by 20%, and the number of netizens is adjusted upward by 10%, which is denoted by the letters A~E, respectively. The simulation results of the effects of advertising promotion, word-of-mouth communication, and the number of netizens on the growth rate of the number of takeout users and the number of takeout users are shown in Figures 13~14, respectively.

From the simulation results in Figure 13, it can be seen that under the original conditions of no change, individually adjusting upward or downward the values of the advertising promotion factor and word-of-mouth communication factor will have a certain impact on the evolution of the number of takeout users. Among them, the influence of the number of Internet users on the evolution of the number of takeout users is the largest, and the increasing popularity of the Internet is an important condition for the development of the takeout O2O industry.

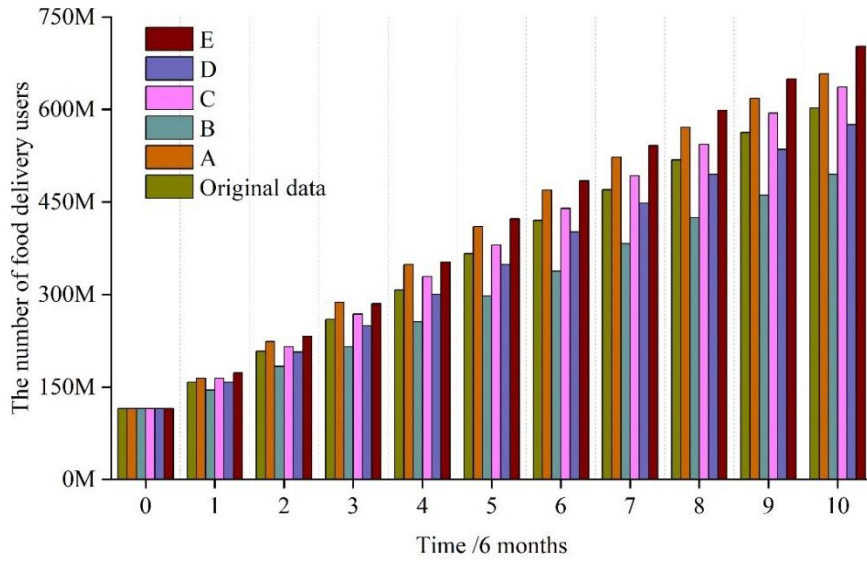


Figure 13: The impact on the number of takeout users

According to the simulation results in Figure 14, the degree of influence of advertisement promotion on the evolution of the number of takeaway users is higher than that of word-of-mouth promotion, indicating that the effect of advertisement promotion in the takeaway O2O industry in the era of “Internet+” by placing advertisements in the media is better than that of traditional interpersonal communication in a short period of time. In addition, the effect of the change of the above variables on the growth rate of the number of takeaway users is phased, and their growth rate curves all show a downward trend after November 2022. This indicates that the number of takeaway users is bound to grow in the short term, but in the long term, limited by conditions such as the degree of Internet penetration, the decline in its growth rate is inevitable. In addition, curve A, after reaching the peak, declines significantly faster than curve C, which also shows that in the era of traffic, the heat is generated quickly, but it also declines quickly, and if you can't continue to make additional investment in advertising and promotion, you need to combine the word-of-mouth communication between people to slow down the decline of the heat of the promotion of the takeaway O2O model.

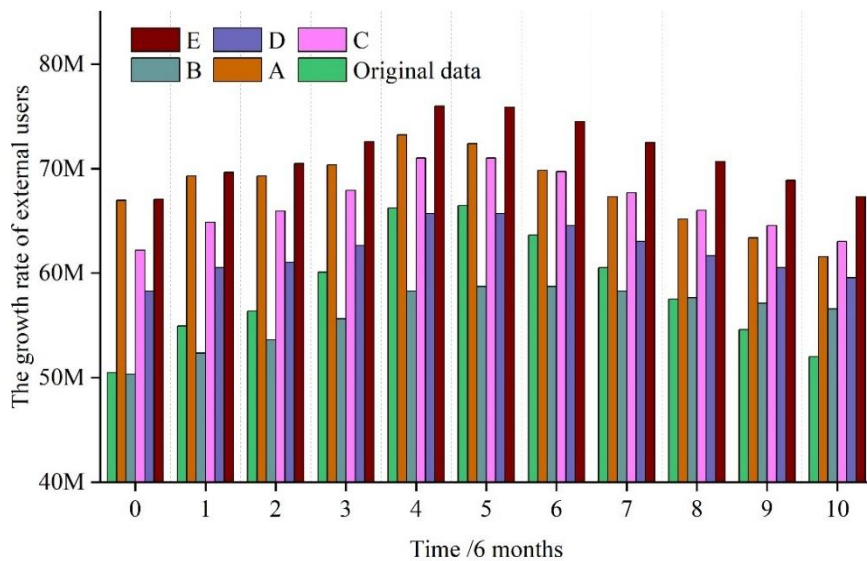


Figure 14: The impact on the growth rate of takeout users

Similarly, the simulation of takeaway market size, logistics capacity and its growth rate can lead to the following conclusions:

(1) Takeaway market size is positively affected by product quality and logistics quality, but product quality has a greater degree of influence, and only if product quality is guaranteed, the corresponding improvement of logistics service quality will be conducive to the development of takeaway O2O industry.

(2) The number of merchants and types of meals also have a greater positive impact on takeaway market size.

(3) Changes in consumption level and takeaway pricing have a great impact on the evolution of takeaway market size. Rising consumption level can promote the increase of takeaway market size. At the same time, consumers' sensitivity to takeaway meal pricing is high, and a reasonable subsidy strategy can help expand the takeaway market size, while too high pricing can limit the development speed of the takeaway industry.

(4) The evolutionary trend of logistics capacity mainly depends on the demand for logistics capacity in the takeaway industry. There is a certain gap between the logistics capability of the logistics distributor and the takeaway industry's demand for logistics capability, which prompts the logistics distributor to invest in the logistics business and ultimately improves the logistics distributor's own distribution capability.

4 Conclusion

In this paper, an LSTM-Transformer-based method for predicting the evolutionary trend of business ecosystems is proposed, which solves the problems of low accuracy and low robustness of existing prediction methods.

The LSTM-Transformer model has a small deviation between the predicted and actual values of total customer consumption demand, and can fit the takeaway platform data well. In single-step prediction, the LSTM-Transformer has the highest prediction accuracy, corresponding to the smallest MAE, MAPE, and SMAPE values, which are 4.124 yuan, 20.316%, and 22.057%, respectively, relative to the models of CV-KF, LSTM, LSTM-Attention, and Transformer. And the model also achieves the optimal prediction in multi-step prediction, which indicates a clear advantage in LSTM-Transformer long sequence trajectory prediction. In addition, the LSTM-Transformer model learns these temporal dependencies directly from the input data, suppressing irrelevant variables so that they do not affect the model prediction, and this ability ensures the credibility of the model.

Based on the simulation results of the LSTM-Transformer model for the takeout O2O business ecosystem, this paper puts forward the following management recommendations:

(1) The evolution of the number of takeout users is influenced by advertising promotion, word-of-mouth promotion, and the number of Internet users. It is possible to increase the investment in advertising and promotion in the media to realize user diversion, and rationally allocate diversion resources by combining the advantages of advertising and promotion and word-of-mouth promotion, and at the same time, according to the total population and the degree of Internet penetration, formulate the corresponding development strategy for different regions to avoid the waste of costs caused by blind promotion.

(2) With regard to the evolution of the takeaway market scale, takeaway platforms should do a good job of checking product quality, and do a good job of guaranteeing the security of the platform's transportation and payment, and at the same time, improve the quality of the logistics service of the takeaway industry, and formulate a series of measures to improve the enthusiasm of catering merchants in joining the takeaway platforms.

(3) Government departments need to increase support for the catering industry to ensure the

operation and development of catering merchants. Merchants should also actively comply with the development of the takeaway O2O industry, adjust their business thinking in time, increase the investment in the takeaway business, and actively develop new dishes.

Funding

This work is supported by the Youth Key Project of Humanities and Social Sciences in Shandong Province in 2025.

About the Author

Na Han was born in Zaozhuang, Shandong in September 1988. She obtained her Bachelor's degree in Management from Shandong University in July 2010 and her Master's degree in Business Administration from Northeast Normal University in July 2012. Currently, she serves as a dedicated lecturer at the School of Economics and Management, Zaozhuang University.

References

- [1] Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International journal of information management*, 47, 52-64.
- [2] Mäntymäki, M., & Salmela, H. (2017). In Search for the Core of the Business Ecosystem Concept: A Conceptual comparison of business ecosystem. In 9th International Workshop on Software Ecosystems (IWSECO 2017) (p. 103).
- [3] Rinkinen, S., & Harmaakorpi, V. (2018). The business ecosystem concept in innovation policy context: Building a conceptual framework. *Innovation: the European journal of social science research*, 31(3), 333-349.
- [4] Wieringa, R. J., Engelsman, W., Gordijn, J., & Ionita, D. (2019, July). A business ecosystem architecture modeling framework. In 2019 IEEE 21st Conference on Business Informatics (CBI) (Vol. 1, pp. 147-156). IEEE.
- [5] Lindgren, P. (2016). The Business Model Eco-System. *Journal of Multi Business Model Innovation and Technology*, 4(2), 1-50.
- [6] Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research policy*, 47(8), 1391-1399.
- [7] Lee, Y. W., Moon, H. C., & Yin, W. (2020). Innovation process in the business ecosystem: the four cooperations practices in the media platform. *Business Process Management Journal*, 26(4), 943-971.
- [8] Peltoniemi, M., Vuori, E., & Laihonon, H. (2005, September). Business ecosystem as a tool for the conceptualisation of the external diversity of an organisation. In *Proceedings of the Complexity, Science and Society Conference* (pp. 11-14). Liverpool, Great Britain.

- [9] Peltola, T., Aarikka-Stenroos, L., Viana, E., & Mäkinen, S. (2016). Value capture in business ecosystems for municipal solid waste management: Comparison between two local environments. *Journal of cleaner production*, 137, 1270-1279.
- [10] Aarikka-Stenroos, L., & Ritala, P. (2017). Network management in the era of ecosystems: Systematic review and management framework. *Industrial marketing management*, 67, 23-36.
- [11] Cennamo, C., & Santaló, J. (2019). Generativity tension and value creation in platform ecosystems. *Organization science*, 30(3), 617-641.
- [12] Burford, S., & Resmini, A. (2017). Cross-channel information architecture for a world exposition. *International Journal of Information Management*, 37(6), 547-552.
- [13] Mäkinen, S. J., Kannianen, J., & Peltola, I. (2014). Investigating adoption of free beta applications in a platform-based business ecosystem. *Journal of Product Innovation Management*, 31(3), 451-465.
- [14] Fujita, M., Okudo, T., & Nagane, H. (2021). Analyzing business ecosystem through corporate networks based on capital relations. *Procedia CIRP*, 99, 692-697.
- [15] Cui, Z., Ouyang, T., Chen, J., & Li, C. (2019). From peripheral to core: A case study of a 3D printing firm on business ecosystems reconstruction. *Technology Analysis & Strategic Management*, 31(12), 1381-1394.
- [16] Aksoy, C. (2023). Digital business ecosystems: An environment of collaboration, innovation, and value creation in the digital age. *Journal of business and trade*, 4(2), 156-180.
- [17] Hein, A., Weking, J., Schrieck, M., Wiesche, M., Böhm, M., & Krcmar, H. (2019). Value co-creation practices in business-to-business platform ecosystems. *Electronic markets*, 29(3), 503-518.
- [18] Saeedi, R. (2023). Ecological value creation: A novel economic paradigm of interactions in business ecosystems. *World Futures*, 79(2), 320-346.
- [19] Pera, R., Occhiocupo, N., & Clarke, J. (2016). Motives and resources for value co-creation in a multi-stakeholder ecosystem: A managerial perspective. *Journal of business research*, 69(10), 4033-4041.
- [20] Rong, K., Lin, Y., Li, B., Burström, T., Butel, L., & Yu, J. (2018). Business ecosystem research agenda: More dynamic, more embedded, and more internationalized. *Asian Business & Management*, 17(3), 167-182.
- [21] Nuutinen, M., Valkokari, K., & Jurmu, M. (2025). Critical features of ecosystem practice empowering industrial sustainability transformation. *Creativity and Innovation Management*, 34(1), 111-124.
- [22] Kumar, P., Sharma, A., & Salo, J. (2019). A bibliometric analysis of extended key account management literature. *Industrial Marketing Management*, 82, 276-292.

- [23] Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. *Journal of management*, 43(1), 39-58.
- [24] Clarysse, B., Wright, M., Bruneel, J., & Mahajan, A. (2014). Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. *Research policy*, 43(7), 1164-1176.
- [25] Awano, H., & Tsujimoto, M. (2021). The mechanisms for business ecosystem members to capture part of a business ecosystem's joint created value. *Sustainability*, 13(8), 4573.
- [26] Alet, J. (2023). Key pillars in a business ecosystem success. *Journal of Business Strategy*, 44(5), 294-300.
- [27] Felch, V., & Sucky, E. (2023). In search of a consensus definition of business ecosystems: a qualitative study. *Journal of Modelling in Management*, 18(6), 1834-1857.
- [28] Mäkinen, S. J., & Dedehayir, O. (2014). Business ecosystems' evolution—An ecosystem clockspeed perspective. In *Collaboration and competition in business ecosystems* (Vol. 30, pp. 99-125). Emerald Group Publishing Limited.
- [29] Gao, R., Zhang, Z., Tian, Z., Guan, X., & Li, A. (2013). Modelling the emergence and evolution of e-business ecosystems from a network perspective. *Studies in Informatics and Control*, 22(4), 339-348.
- [30] Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic management journal*, 39(8), 2255-2276.
- [31] Bertassini, A. C., Zanon, L. G., Azarias, J. G., Gerolamo, M. C., & Ometto, A. R. (2021). Circular Business Ecosystem Innovation: A guide for mapping stakeholders, capturing values, and finding new opportunities. *Sustainable Production and Consumption*, 27, 436-448.
- [32] Yoon, C., Moon, S., & Lee, H. (2022). Symbiotic relationships in business ecosystem: A systematic literature review. *Sustainability*, 14(4), 2252.
- [33] Kapoor, R., & Lee, J. M. (2013). Coordinating and competing in ecosystems: How organizational forms shape new technology investments. *Strategic management journal*, 34(3), 274-296.