


The Multimodal Emotion Information Analysis of E-Commerce Online Pricing in Electronic Word of Mouth

Jinyu Chen, Southwest University of Science and Technology, China

Ziqi Zhong, London School of Economics and Political Science, UK

 <https://orcid.org/0000-0002-3919-9999>

Qindi Feng, Gachon University, South Korea*

Lei Liu, Chuzhou University, China

ABSTRACT

E-commerce has developed rapidly, and product promotion refers to how e-commerce promotes consumers' consumption activities. The demand and computational complexity in the decision-making process are urgent problems to be solved to optimize dynamic pricing decisions of the e-commerce product lines. Therefore, a Q-learning algorithm model based on the neural network is proposed on the premise of multimodal emotion information recognition and analysis, and the dynamic pricing problem of the product line is studied. The results show that a multi-modal fusion model is established through the multi-modal fusion of speech emotion recognition and image emotion recognition to classify consumers' emotions. Then, they are used as auxiliary materials for understanding and analyzing the market demand. The long short-term memory (LSTM) classifier performs excellent image feature extraction. The accuracy rate is 3.92%-6.74% higher than that of other similar classifiers, and the accuracy rate of the image single-feature optimal model is 9.32% higher than that of the speech single-feature model.

KEYWORDS

Dynamic Pricing, E-Commerce, Emotion Recognition, Neural Network, Q-Learning Algorithm

1. INTRODUCTION

1.1 Research Background and Motivations

The development of science and technology is constantly changing people's daily habits. E-commerce has become indispensable in people's lives after its development in recent years (Carta et al., 2018). E-commerce is increasingly becoming the primary tool for selling goods to the mass market (Li et al.,

DOI: 10.4018/JGIM.315322

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

2020). This has led to a growing interest in algorithms and techniques that can predict the future price of a product. They can build intelligent systems that provide affordable goods and services for people (Guan et al., 2020). Online e-commerce platforms provide sellers and consumers with convenient and inexpensive means to facilitate information sharing and docking trade-offs (Wu et al., 2020). The platform economy and e-commerce have become the fastest growing and most dynamic economic sectors in China and the world (Xu et al., 2021). Market competition is becoming increasingly fierce, and consumers' purchasing choices are closely related to product brands. The positive role of brands is increasingly prominent (Leung et al., 2019). An online shopping platform is a unique network group organization that connects products and consumers' purchasing channels. It has gradually become an important communication channel for online brand product sales, profoundly affecting consumers' purchasing choices and brand building (Chen et al., 2020). Dynamic pricing can promote consumption, and the research on dynamic pricing of e-commerce products is necessary.

1.2 Research Objectives

The pricing strategy of e-commerce product lines must be in line with consumers' demand and psychology to stand out in increasingly fierce market competition (Amanah and Harahap, 2018). Dynamic adjustment of prices in different scenarios can stimulate consumers to consume. However, the product's reputation should be considered to promote consumer consumption and enhance their interest in long-term benefits. Therefore, the pricing of e-commerce product lines is critical. Besides, research on pricing strategies is a hot topic. The e-commerce pricing strategy is studied based on multimodal emotion recognition to optimize the pricing strategy of e-commerce. Compared with traditional e-commerce pricing, the proposed method dynamically prices products by studying the influence of consumers' emotions on their consumption behaviors.

2. LITERATURE REVIEW

At present, many scholars have studied the dynamic pricing of e-commerce products. Tseng et al. (2018) stated that online emotion analysis could be used to explore various possibilities, ranging from effects on product prices to impact on sales behavior and essential brand strategies (Tseng et al., 2018). Lv et al. (2020) argued that price discounts could significantly increase social e-commerce sales due to online consumer interactions (Lv et al. 2020). Liu et al. (2022) claimed that consumers increasingly demanded diversified online services. Many online service providers competed fiercely in providing online composite services to meet market demands (Liu et al., 2022). Moerth-Teo et al. (2021) pointed out that it is necessary to scientifically guide e-commerce to price products and improve the profits of service providers to meet consumers' most extraordinary service demands (Moerth-Teo et al., 2021). Ali and Bhasin (2019) indicated that consumers' perceptions of price and delivery quality greatly impacted subsequent shopping intentions. Therefore, the right price greatly influenced consumers' repurchasing (Ali and Bhasin, 2019). Keller et al. (2022) studied the impact of price discounts on dynamic pricing. The results showed that dynamic pricing reduced consumers' willingness to buy, but a large discount could reduce negative consumer sentiment (Keller et al., 2022). Kayikci et al. (2022) proposed a dynamic pricing model. The model used real-time Internet of Things sensor data to contribute significantly to merchants' dynamic pricing at different stages of the sales season (Kayikci et al., 2022).

To sum up, the pricing strategy of e-commerce greatly affects consumers' purchasing choices and the brand reputation of products. Many factors need to be considered in pricing, which should be scientifically sound. The research on dynamic pricing of the product line has important research value.

3. RESEARCH METHODOLOGY

3.1 Multimodal Emotion Recognition

Emotion recognition has developed into a mature discipline in machine learning and artificial intelligence (He et al., 2022). Commonly used methods for multimodal emotion recognition technology are Naive Bayes, decision tree learning, maximum entropy, Support Vector Machines (SVM), and Random Forest (RF) (Abdullah et al., 2021). Emotion recognition is mainly studied from speech and image modalities.

The video data is divided into speech and image data, and the emotional features of the two are extracted respectively to obtain the expression recognition model of a single feature. Then, the two models are fused to obtain a multimodal integrated expression recognition model (Siddiqui et al., 2022). Figure 1 shows the specific process.

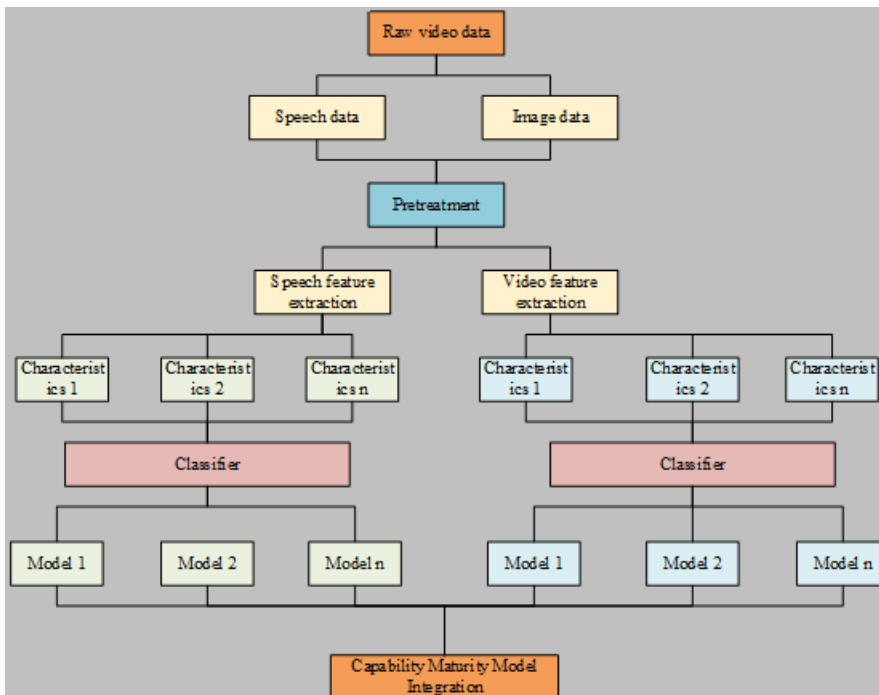
First, the raw video data is divided into speech and image data, and the emotional features of the speech and image data are extracted. Then, the obtained speech and image emotion features are classified through a classifier to obtain several emotion models. Finally, the speech model and the image model are combined for decision-making to obtain an integrated model.

3.1.1 Speech Feature Extraction

The main features of speech modality are divided into Mel Frequency Cepstrum Coefficient (MFCC) features and features extracted by the SoundNET Convolutional Neural Network (CNN). It also includes frame-level features such as IS09, IS11, and IS13. Besides, frame-level features such as IS09, IS11, and IS13 are extracted using specific tools (Cimtay et al., 2020).

The extraction of MFCC first performs Fast Fourier Transform (FFT) on the sampled speech sequence $x(n)$. The equation for FFT is as follows.

Figure 1. Test flow chart



$$X(k) = \sum_{n=0}^{N-2} x(n) e^{-j\frac{2\pi}{N}nk}, k = 0, 1, 2, \dots, N-1 \quad (1)$$

In Eq. (1), N represents the frame length.

Next, the Mel filter is configured. In addition, the filtered output is calculated. The frequency response $H_m(k)$ corresponding to the filter is:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{2(k - f(m-1))}{(f(m+1) - f(m-1))(f(m) - f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1) - k)}{(f(m+1) - f(m-1))(f(m) - f(m-1))}, & k < f(m-1) \\ 0, & k \geq f(m+1) \end{cases} \quad (2)$$

In Eq. (2), the center frequency corresponding to the filter is $f(m)$.

Then, the filter bank output's logarithmic energy $S(m)$ is calculated.

$$S(m) \ln \left(\sum_{K=0}^{N-1} |X_a(k)|^2 H_m(k) \right), 0 \leq m \leq M \quad (3)$$

In Eq. (3), the number of filters is represented by M .

Finally, the corresponding coefficient $C(n)$ can be obtained through discrete cosine change.

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos \left(\frac{\pi n(m - 0.5)}{M} \right), n = 1, 2, \dots, L \quad (4)$$

In Eq. (4), the order corresponding to the MFCC coefficient is L .

The main advantage of the SoundNet CNN is that it has good learning ability and can realize the extraction of speech information. The corresponding realization principle is as follows.

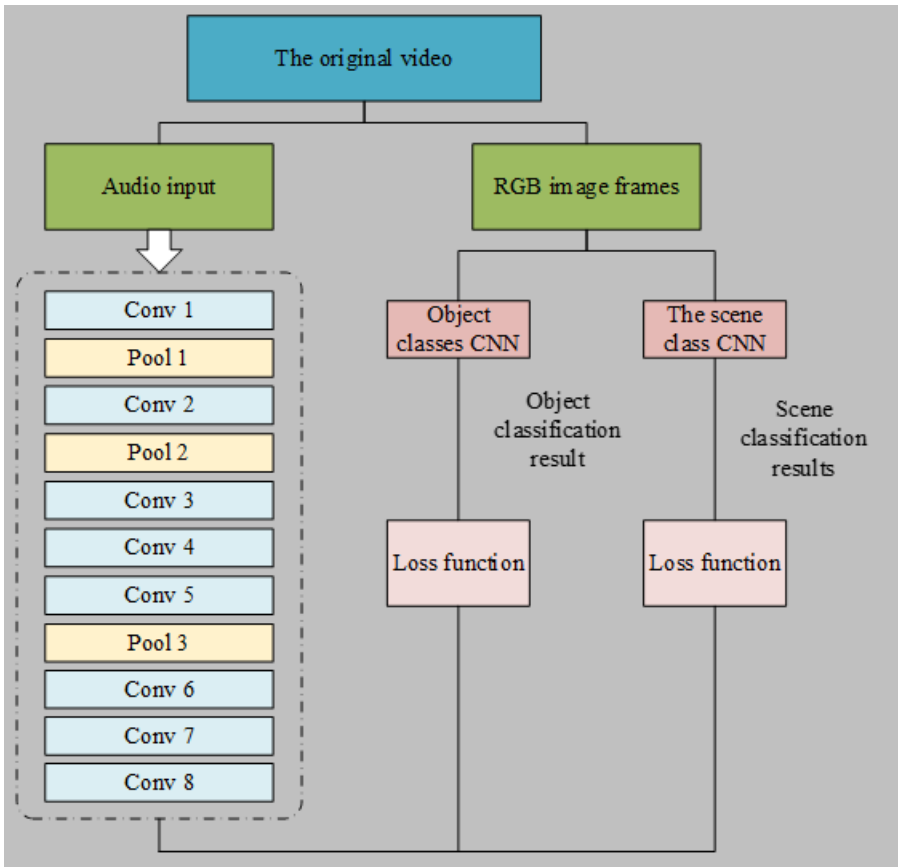
First, the video is cut into audio and RGB (Red Green Blue) image frames. The RGB image frame part is identified and classified using two kinds of NNs: image and scene NNs. The structure diagram of the SoundNet network is shown in Figure 2.

Figure 2 displays the schematic diagram of the structure of the SoundNet network. Conv n in the figure represents the n^{th} convolutional layer, and Pool n represents the n^{th} pooling layer.

3.1.2 Image Feature Recognition

The features of the imaging modality refer to the features extracted by the CNN using DenseNet (Dense Connection) and Visual Geometry Group (VGG) and the features obtained by the feature scan sub (Siriwardhana et al., 2020).

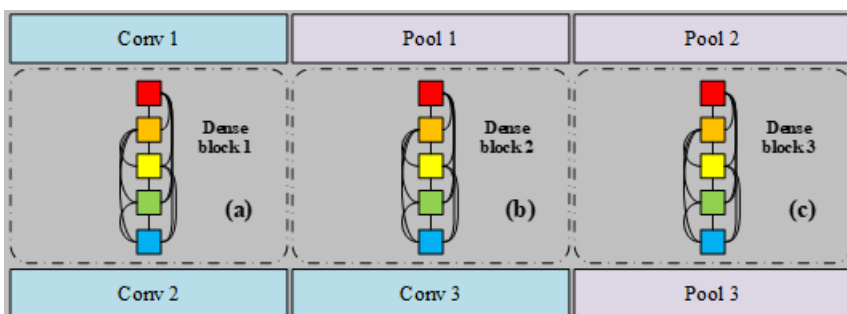
Figure 2. Structure of the SoundNet network



DenseNet network has excellent performance (Tong et al., 2020). A particular DenseNet-BC network is used in the feature extraction process. The corresponding structure diagram is shown in Figure 3.

From Figure 3, the structure of the DenseNet-BC network is divided into two parts. One is the bottleneck layer, and the other is the transition layer. Pooling and convolutional layers can be combined to form a corresponding transition layer.

Figure 3. DenseNet-BC network structure diagram(a) Dense block 1; (b) Dense block 2; (c) Dense block 3



3.1.3 Selection of Classifiers

Various classic classifiers, including SVM and RF, are selected (Zhang, 2020). The Long Short-Term Memory (LSTM) of the long sequence of image frames obtained after classification has apparent advantages in the processing process (Nemati et al., 2019). So, an LSTM-based classifier is designed. Figure 4 displays its structure.

Figure 4 indicates that the input sequence X is the features at any time, and they differ. Multiple LSTM nodes form a new LSTM array to obtain feature information after a batch normalization layer is added after the input layer (Nguyen et al., 2018).

3.1.4 Fusion of the Decision Layers

Common decision fusion methods in multimodal emotion recognition include weighted voting and average methods (Wu et al., 2022). Each of the above models is independent during training, so the weighted voting method is used to fuse the decision layers (Huang et al., 2019). The specific process is as follows.

It is assumed that M represents the number of expression categories, L means the specific number of models, the i^{th} emotion recognition model is represented by h_i , its corresponding weight is w_i , and w_i meets the following conditions.

$$\begin{cases} \sum_{i=1}^L w_i = 1 \\ w_i \in [0, 1], i = 1, 2, \dots, L \end{cases} \quad (5)$$

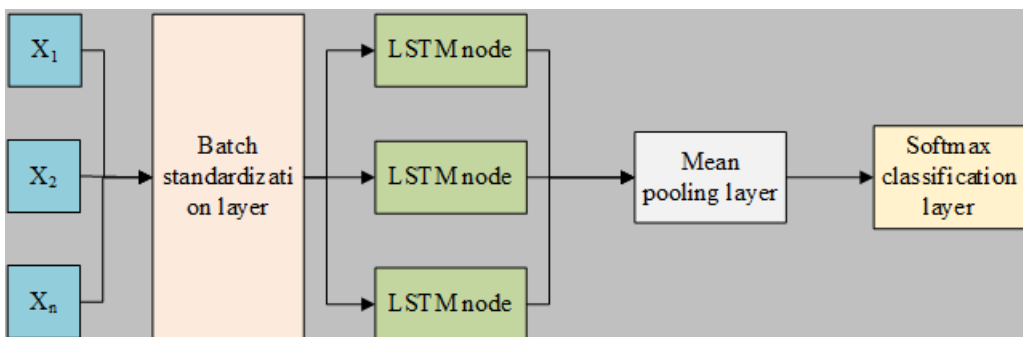
It is assumed that $f(x)$ is the set of weighted voting values corresponding to each expression category obtained using the weighted voting method, and $y(x)$ represents the decision result corresponding to the expression category, and then:

$$f_j(x) = \sum_{i=1}^L w_i I(h_i(x), j) \quad (6)$$

$$y(x) = \operatorname{argmax}_j f_j(x) \quad (7)$$

In Eq. (6)-Eq. (7), $j = 1, 2, \dots, M$. The corresponding definition of function I is as follows.

Figure 4. Structure diagram of LSTM classifier



$$I(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (8)$$

3.2 Dynamic Product Pricing Decision Model

The dynamic pricing decision of products adopts Markov Decision Process (MDP). The entire decision must consider the current state without considering the historical state (Xie et al., 2018). Partially Observable Markov Decision Process (POMDP) means that decision-makers must make decisions using partially observed information in random scenarios (Chen et al., 2020). Since the current state cannot be directly observed during the entire observation process, or there is a lack of information during the observation process, it is necessary to make decisions based on some existing information (Gunawan et al., 2019). The dynamic pricing decision of products is a problem of POMDP optimization. In dealing with this problem, e-commerce-related systems can only obtain the current product inventory, delivery volume, and product sales volume (Sullivan and Kim, 2018). Information about the refined proportional structure of consumers at present is not available. The Bellman recurrence equation corresponding to the product line revenue under the dynamic pricing strategy based on the POMDP problem model is proposed to solve the above problems (Wang et al., 2020).

Assuming that the residual value of the commodity at the end of the cycle is zero, the inventory of commodity w at the beginning of period t is denoted by I_t^w .

The inventory vector and sales vector are denoted by I_t and d_t , respectively. Therefore, the corresponding inventory vector in the $t+1$ period can be obtained as $I_{t+1} = I_t - d_t$. The set of sales vectors is denoted by D_t . It is assumed that the demand structure in the current state is g , the price vector corresponding to the product line is P_t , and $R_t(I_t, \lambda_t, \psi_t)$ represents the revenue from the period t to the end of the entire sales process. The Bellman recurrence equation of the product line revenue function under the dynamic pricing strategy is:

$$R_t(I_t, \lambda_t, \psi_t) = \max_{P_t} \left(\sum_{g=1}^G \sum_{d_t} \psi_t^g Pr_{purchase}(d_t | g, \lambda_t N, P_t) \left(\sum_{w=1}^W P_t^w d_t^w + R_{t+1}(I_t - d_t, \lambda_{t+1}, \psi_{t+1}) \right) \right) \quad (9)$$

3.3 Q-Learning Algorithm Based on the Simulation Object Model (SOM)

3.3.1 The Basic Principle of Q-learning

Assuming that s is the state of a scene and d represents a decision, the quick-check table of the training results of Q-learning is shown in Table 1.

Table 1. Q Quick reference of learning and training results

State s	s_1	s_2	s_3	s_t
Decision d	d_1	d_2	d_3	d_t
Q value	Q_1	Q_2	Q_3	Q_t

The system does not need to perform calculations when it completes a given decision-making task. The decision d that can maximize the Q value can be queried from the above table through the current state s . Therefore, the Q algorithm can solve the real-time needs of decision-making.

In the simulation learning process, the intelligent agent can execute the corresponding decision d' by observing the state s_t at any time t . After accepting the decision, the environment will implement a state transition of $s_t \rightarrow s_{t+1}$ and also receive a real-time reward $f(s_t, d')$. Then, the $Q(s_t, d')$ value is simply adjusted by the following equation.

$$Q(s_t, d') = (1 - \beta)Q(s_t, d') + \beta(f(s_t, d') + \gamma V_{(s_{t+1})}) \quad (10)$$

In Eq. (10), $V_{(s_{t+1})} = \max_d(Q(s_{t+1}, d))$. The learning rate parameter is denoted by β , and the time discount rate corresponding to the gain is represented by γ .

The Q values of various (s, d) combinations will tend to converge after iterative calculations.

3.3.2 Application of Q-learning in SOM Model

The system state that the decision depends on is defined as $s_t = (I_t, \lambda_t, \psi_t, t')$ according to the dependence of the reaction in Eq. (9), and $Nt' = (T - t)$. In practice, merchants usually use discounts to adjust prices. Let $0 < \alpha_1 < \alpha_2 < \dots < \alpha_{m-1} \leq 1$, the system can only select the appropriate price discount rate from the set $\{\alpha_1, \alpha_2, \dots, \alpha_{m-1}, 1\}$. The number of decision options and states in the above problem is enormous, so the cheat sheet cannot satisfy many decision and state spaces.

The main obstacle to the practical application of Q-learning is the vast space problem (Jang et al., 2019). Then, SOM (Simulation Object Model) and Q clustering are integrated using NN's excellent information storage ability.

In the trading simulation in the decision period, a Monte Carlo Simulation (MCS) of the trading behavior in a decision period is carried out to obtain the income data of any period.

The first step is to randomly select the type of demand in the decision period t using the probability distribution ψ_t .

The second step is to simulate customer arrivals in probability λ_t . Meanwhile, the customer segment data that has arrived is determined according to the probability distribution \hat{B}_g .

The third step is to simulate customers' purchase behavior according to the equations.

The network structure of the SOM adopts a single-layer structure. The corresponding input vector states and decisions are $s_t = (I_t, \lambda_t, \psi_t, t')$, and $d = P_t$, respectively. The connection weights corresponding to the nodes and the input vectors are determined as follows.

- (1) The connection weight corresponding to node i and state s_t is $v_{i,I}^1, \dots, v_{i,I}^W, v_{i,\lambda}, v_{i,\lambda}^1, \dots, v_{i,\psi}^G, v_{i,t'}$.
- (2) The connection weight corresponding to node i and decision d is $v_{i,d} = (v_{i,p}^1, \dots, v_{i,p}^W)$.
- (3) Let $v_{i,Q}$ take the utility of the decision d represented by the state s represented by the node i , so $Q(s, d)$.

Q-learning is built on the following principles based on the above SOM structure.

- (1) After any MCS calculation, Q-learning only selectively updates the node weights of the currently calculated state based on the SOM algorithm to avoid overgeneralization.
- (2) The system needs to estimate the current and future environmental parameters through some information.
- (3) The update process of the node weight needs to determine the input vector and SOM similar nodes according to Eq. (11).

$$i = \arg \min_t \left(|s_t - v_{i,s}| + |P_t - v_{i,d}| \right) \quad (11)$$

After that, the weight vector corresponding to the current node is updated using the SOM algorithm. According to the sales revenue obtained by the simulation, Eq. (10) is used to modify the weight $v_{i,Q}$ of the Q value of the current corresponding node.

- (4) The fourth is the principle of dimensional consistency. Let $t' = \frac{t'}{TN}$, $I_w^t = \frac{I_w^t}{I_w^1}$, and $P_t^w = \frac{P_t^w}{P_{max}^w}$.

P_{max}^w represents the most possible price of commodity w in the entire market. For the return f

obtained from any transaction, $f = \frac{f}{R_{max}}$. R_{max} represents the income received when the goods

are sold at the highest price. During the above conversion process, the value interval of the numerical value of each calculation quantity is guaranteed to be in the range of [0, 1].

- (5) The fifth is to choose the optimal strategy. According to the principle that “decisions taken in similar states are also similar,” the optimal strategy is selected from similar strategies of multiple nodes. The selection process can be carried out according to Eq. (12).

$$i = \arg \min_t \left(|s_t - v_{i,s}| + |1 - v_{i,Q}| \right) \quad (12)$$

- (6) The node weight vector is transformed into a decision-making scheme. The specific numerical value represented by the node is converted into a specific decision by Eq. (13).

$$p_t^w = p_{max}^w v_{i,p}^w \quad (13)$$

The dynamic pricing SOM-learning algorithm for product lines designed according to the above principles is shown in Figure 5.

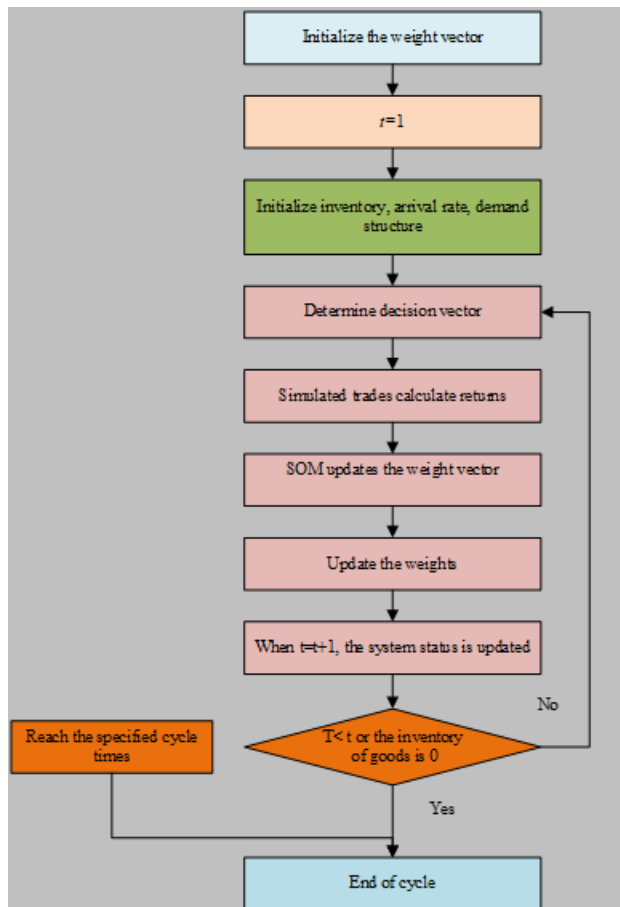
4. EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

4.1 Experimental Environment

4.1.1 Experimental Design of Multimodal Emotion Recognition

The experimental environment is Windows 10 operating system. The multimodal emotion recognition is studied based on Acted Facial Expressions in the Wild (AFEW) dataset. The obtained raw data is preprocessed. The dataset is divided into speech data and image data. Then, the emotional features of speech

Figure 5. Dynamic pricing SOM-learning algorithm for product line



and image are extracted from the two data types. The experimental process is data preprocessing, building an NN model, training, testing, and evaluating the model. The feature extraction method is as follows.

The input data is the audio data obtained after video segmentation, and the extracted features are SoundNet in extracting features using the SoundNet CNN. The DenweNet-BC CNN is used to extract features in the feature extraction for image data. The output result of the last average pooling layer is used as the final feature result, denoted as DenweNet-pooling 3. The 12th convolutional layer and the 1st fully connected layer of the VGG-16 NN are used as the final output result, which is the final feature result, denoted as VGG-conv 13 and VGG-fc1, respectively. The features extracted by the feature descriptor based on Local Binary Patterns from Three Orthogonal Planes are obtained from the AFEW dataset, denoted as LBP-YOP.

4.1.2 Experiment Design based on Product Line Pricing Model

Some market environment parameters of the simulation experiment are based on the market research data of three women's dresses. Dresses are seasonal products, so the research time is chosen in summer. Three women's dress products on e-commerce platforms are randomly selected, and market research is conducted on the three dresses. Then, the data obtained from the survey is analyzed and processed to obtain the market demand structure corresponding to the three dresses. The current demand structure matrix is appropriately processed through data analysis. After a period, each structure type has a tiny probability of

changing to the other three market demand structure types. Let $a_0 = A_i, t$ denote the possibility that the i^{th} demand structure remains unchanged after a period. The probability that $a_1 = A_{i,i+1}, a_2 = A_{i,i+2}, a_3 = A_{i,i+3}$ can be transformed into the other three demand structures, and $a_0 + a_1 + a_2 + a_3 = 1$. MCS is used to find optimal returns for static pricing to verify the method's practicality further. Also, the inventory parameters are adjusted, and three dynamic pricing simulation tests are conducted.

4.2 Parameters Setting

The experimental parameters of the e-commerce product line pricing model are shown in Table 2.

4.3 Performance Evaluation

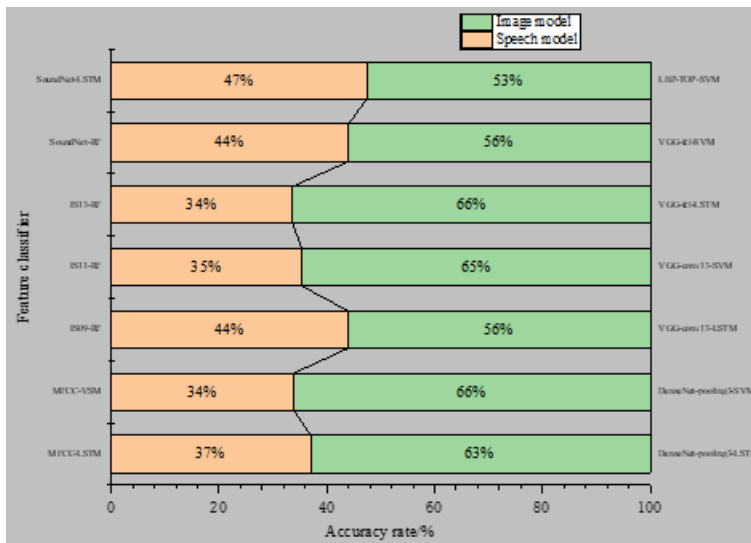
4.3.1 Experimental Results of Multimodal Emotion Recognition

SVM, RF, and LSTM-based classifiers are used to classify emotions after the feature extraction of speech and image models is achieved. Then, multiple single-feature expression recognition models based on speech and images are obtained. Figure 6 demonstrates the classification results based on speech and image single-feature models.

Table 2. Experimental parameters of e-commerce product line pricing model

Experimental parameters	The meaning of the parameter expression	Parameter value
T	It contains T decision periods.	5
N	It contains N periods.	1500
I_1	Initial stock vector	(350, 400, 500)
λ_1	Initial arrival rate state vector	0.6
K	It has K reserved price ranges.	30
C	C customer segments	5
G	The market demand structure category has G .	8
$[v_{min}^1, v_{max}^1]$	Reserve price range corresponding to product 1	[30, 240]
$[v_{min}^2, v_{max}^2]$	Reserve price range corresponding to product 2	[100, 350]
$[v_{min}^3, v_{max}^3]$	Reserve price range corresponding to product 3	[150, 450]
β	Learning rate parameter	0.2
γ	Time discount rate corresponding to income	0.99
ψ_1	Initial demand structure state vector	(0.40, 0.35, 0.30, 0, ..., 0)
a_0, a_1, a_2, a_3	Probability of demand structure shift	(0.15, 0.35, 0.30, 0.25)

Figure 6. Recognition accuracy of single feature expression



It is found that the accuracy of the LSTM classifier in terms of speech traits has an improvement compared with that of SVM and RF by analyzing the information in Figure 6(a). The model based on IS09 features has the highest accuracy of 33.56% in speech recognition. From Figure 6(b), the LSTM classifier's accuracy in the image feature classification process is 3.92%-6.74% higher than that of other classifiers. Furthermore, the model based on VGG-conv13 features has the highest accuracy of 42.88%. Compared with the optimal model of a single feature of speech, the optimal model of a single feature of the image has significantly higher accuracy, 9.32% higher than that of speech classification.

Three fusion models are obtained using the weighted voting method to fuse multiple single-feature models. The classification results based on the fusion model are shown in Table 3.

From Table 3, the classification accuracy of the speech and image dual-modal fusion model reaches 50.62%, which is higher than that of all single-modal models.

Figure 7 shows the official AFEW dataset of the multimodal emotion information recognition EmotiW competition.

From Figure 7, the average accuracy of the contestants' emotion recognition reaches about 50%, and the highest accuracy reaches 62.3%. Although it is lower than the existing highest standard, the accuracy can maintain at a high level, which has application value.

4.3.2 Experimental Results of Dynamic E-commerce Pricing

In the case of different SOM nodes, the benefits obtained through dynamic pricing and the computational cost of the simulation experiment are shown in Figure 8.

From Figure 8(a), the model system has an obvious over-generalization problem when the number of nodes is in the interval of [100, 250], which leads to the benefit of dynamic pricing being much

Table 3. Accuracy of the Fusion Model

Fusion model	Image modality fusion	Speech modal fusion	Dual-modal fusion model
Added single feature model	3, 5, 6, 10, 12	1, 4, 6, 8	1, 5, 7, 8, 10, 13
Accuracy /%	37.49	47.68	50.62

Figure 7. Audio and video emotion recognition data

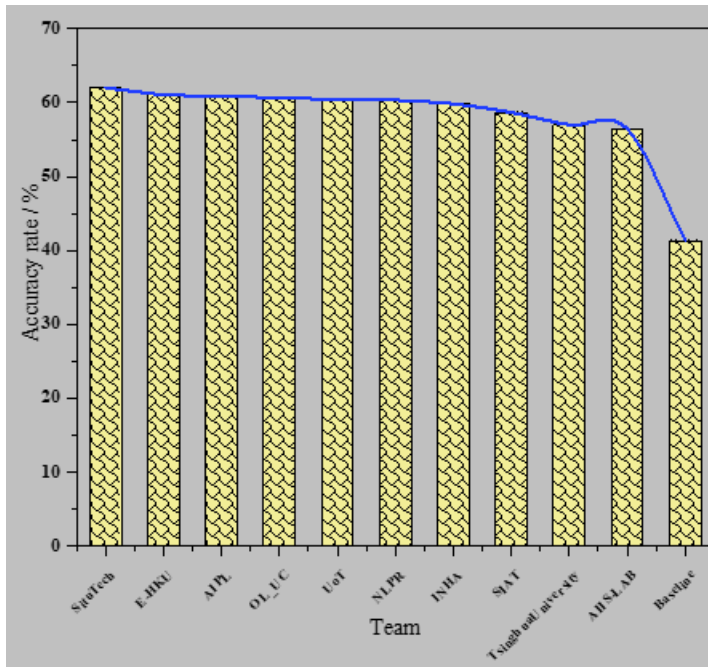
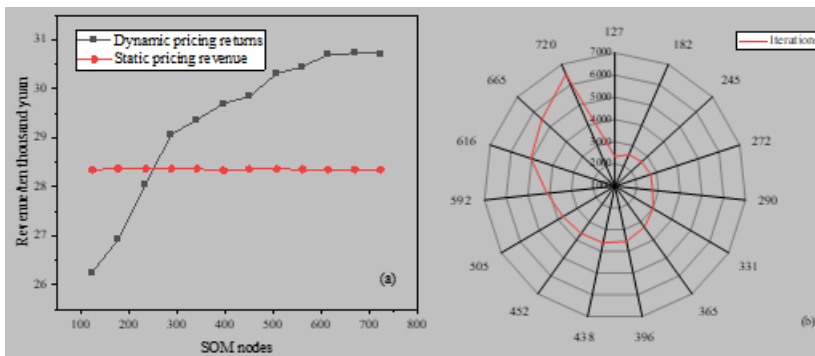


Figure 8. The benefits of dynamic pricing and the computational costs of simulation experiments (a) Dynamic and static revenue comparison; (b) Comparison of calculation amount of different nodes



smaller than that of static pricing. Dynamic pricing tends to be stable when the number of nodes exceeds 500. This shows that state aggregation has little effect on decision-making accuracy under moderate circumstances. Figure 8(b) indicates that the computational cost of the whole experiment increases continuously as the number of nodes in the SOM model continues to grow. The results show that this algorithm needs to properly control the number of nodes in the SOM model based on ensuring the accuracy of decision-making to reduce the computational cost of the algorithm. Table 4 compares the benefits of dynamic and static pricing when the inventory at the beginning of the sales period is different in other markets under constant parameters.

From Table 4, the dynamic pricing obtains a slight change in the income growth rate when the supply and demand can maintain a relative balance. The reason for this phenomenon is that the

Table 4. Dynamic and static revenue comparison

	Beginning inventory vector	Dynamic pricing revenue growth rate/%	Ending remaining inventory vector	
			Dynamic pricing	Static pricing
Experiment 1	(400,450,500)	3.68	(0,0,0)	(0,0,0)
Experiment 2	(300,350,450)	6.75	(9,12,15)	(0,0,0)
Experiment 3	(530,650,750)	14.32	(93,68,41)	(21,13,8)

primary source of the increase in revenue is the surplus of consumers. The revenue growth value brought by dynamic pricing is mainly the surplus of consumers when there is a surplus of inventory goods at the end of the sales period. The dynamic pricing strategy reduces the loss caused by the remaining inventory at the end of the period, so the revenue growth rate of the dynamic pricing strategy is relatively large.

4.4 Discussion

Dynamic pricing is studied based on multimodal emotion recognition to make correct decisions in the pricing strategy of e-commerce product lines. Identifying and classifying consumers' emotions can capture consumers' psychology in pricing. When consumers have consumption behaviors, pricing discounts can be offered to old customers, enhancing the brand's reputation. A comparative analysis of single-modal and multimodal fusion models is carried out in the emotion recognition process. It is found that the multimodal fusion model shows higher performance in recognition accuracy. The research on e-commerce product line pricing finds that dynamic pricing is better than static pricing in revenue, indicating that the study has application value. At present, the research on e-commerce pricing strategy is mostly reflected in consumers' active evaluation and promotion activities. Duan et al. (2022) studied the impact of consumer online reviews and coupons on online product sales and prices (Duan et al., 2022). However, consumers are much willing to evaluate only when the shopping experience is very poor, so it is difficult to reflect the impact of product pricing on consumers' shopping behavior in many cases. The innovation of this paper is to study the consumer's speech and image emotion, which can help to accurately analyze the impact of product pricing on consumers' shopping intention, make targeted pricing to promote consumer behavior, and improve the consumer shopping experience. For example, if a type of product is analyzed through sentiment analysis, and consumers are satisfied with a price point or preferential strategy of a brand of this type of product, the product's pricing can be used as a reference. If they show dissatisfaction, the pricing can be adjusted.

5. CONCLUSION

5.1 Research Contribution

This paper proposes a Q-learning algorithm based on the SOM model for the huge space problem in the dynamic pricing of the product line. The advantage of this algorithm is that the calculation of the SOM model can be integrated with the model training process of the Q-learning algorithm, which provides a highly feasible decision-making optimization method for product line pricing. In addition, pricing is based on multimodal emotion recognition to promote consumer consumption and improve product reputation. Emotions are identified and classified by establishing two modal fusion models of speech and image, and the fusion model is used as an aid to analyze market demand. The results show that the LSTM-based multimodal emotion recognition model exhibits high performance. The accuracy of LSTM classifier in the process of image feature classification is 3.92%-6.74% higher than that of other classifiers. The accuracy of the image single-feature optimal model is 9.32% higher

than that of the speech single-model classification. The multimodal emotion recognition model based on LSTM has high accuracy. The classification accuracy of the audio and image dual-modal fusion model reaches 50.62%, which is higher than that of all single-modal models. It is found that dynamic pricing is more profitable than static pricing when stockists have a surplus after the end of the sales period, compared with dynamic and static pricing. It fully proves the feasibility of the above method, which has high research and uses value.

5.2 Future Works and Research Limitations

However, the expression of emotion is divided into three parts from the aspect of emotion recognition, namely the beginning, the climax, and the end. The recognition of emotion is mainly for the climax part. There are too much redundant data in the speech data research process, which impacts recognition accuracy. It is hoped that in the follow-up research, practical research on emotion recognition is conducted based on more modal fusion through the strategy of feature layer fusion. In terms of pricing strategy for new products, there is a lack of historical data. The pre-knowledge of product pricing cannot be obtained through market research. The impact of the current product on similar products in the entire market can only be analyzed during the development of the product. The relevant market knowledge gained during the product development process can be used to price the product dynamically. Combining up-front knowledge of new products with dynamic pricing is a topic worthy of research and discussion in the coming period.

FUNDINGS

This work was supported by Research on Supply and Demand Coordination of Green Agricultural Products in Sichuan under the Application of Intelligent Agricultural Technology Sichuan Provincial Science and Technology Department (Grant no. 2022JDR0164) and Research on the Ecological Development Model of Smart Agriculture from the Perspective of Ecological Ethics Sichuan Circular Economy Research Center (Grant no. XHJJ-2007).

REFERENCES

- Abdullah, S. M. S. A., Ameen, S. Y. A., & Sadeeq, M. A. (2021). Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(2), 52–58. doi:10.38094/jastt20291
- Ali, A., & Bhasin, J. (2019). Understanding customer repurchase intention in e-commerce: Role of perceived price, delivery quality, and perceived value. *Jindal Journal of Business Research*, 8(2), 142–157. doi:10.1177/2278682119850275
- Amanah, D., & Harahap, D. A. (2018). Examining the effect of product assortment and price discount toward online purchase decision of university student in Indonesia. *Jurnal Manajemen Dan Kewirausahaan*, 20(2), 99–104. doi:10.9744/jmk.20.2.99-104
- Carta, S., Medda, A., Pili, A., Reforgiato Recupero, D., & Saia, R. (2018). Forecasting e-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and google trends data. *Future Internet*, 11(1), 5. doi:10.3390/fi11010005
- Chen, J., Tournois, N., & Fu, Q. (2020). Price and its forecasting of Chinese cross-border E-commerce. *Journal of Business and Industrial Marketing*, 35(10), 1605–1618. doi:10.1108/JBIM-01-2019-0017
- Chen, P., Zhao, R., Yan, Y., & Li, X. (2020). Promotional pricing and online business model choice in the presence of retail competition. *Omega*, 94, 102085. doi:10.1016/j.omega.2019.07.001
- Cimtay, Y., Ekmekcioglu, E., & Caglar-Ozhan, S. (2020). Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access: Practical Innovations, Open Solutions*, 8, 168865–168878. doi:10.1109/ACCESS.2020.3023871
- Duan, Y., Liu, T., & Mao, Z. (2022). How online reviews and coupons affect sales and pricing: An empirical study based on e-commerce platform. *Journal of Retailing and Consumer Services*, 65, 102846. doi:10.1016/j.jretconser.2021.102846
- Guan, X., Deng, W. J., Jiang, Z. Z., & Huang, M. (2020). Pricing and advertising for reward-based crowdfunding products in E-commerce. *Decision Support Systems*, 131, 113231. doi:10.1016/j.dss.2019.113231
- Gunawan, A. V., Linawati, L., Pranandito, D., & Kartono, R. (2019). The Determinant Factors of E-Commerce Purchase Decision in Jakarta and Tangerang. *Binus Business Review*, 10(1), 21–29. doi:10.21512/bbr.v10i1.5379
- He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain–computer interfaces. *Brain Sciences*, 10(10), 687. doi:10.3390/brainsci10100687 PMID:33003397
- Huang, H., Hu, Z., Wang, W., & Wu, M. (2019). Multimodal emotion recognition based on ensemble convolutional neural network. *IEEE Access: Practical Innovations, Open Solutions*, 8, 3265–3271. doi:10.1109/ACCESS.2019.2962085
- Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE Access: Practical Innovations, Open Solutions*, 7, 133653–133667. doi:10.1109/ACCESS.2019.2941229
- Kayikci, Y., Demir, S., Mangla, S. K., Subramanian, N., & Koc, B. (2022). Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers. *Journal of Cleaner Production*, 344, 131068. doi:10.1016/j.jclepro.2022.131068
- Keller, A., Vogelsang, M., & Totzek, D. (2022). How displaying price discounts can mitigate negative customer reactions to dynamic pricing. *Journal of Business Research*, 148, 277–291. doi:10.1016/j.jbusres.2022.04.027
- Leung, K. H., Luk, C. C., Choy, K. L., Lam, H. Y., & Lee, C. K. M. (2019). A B2B flexible pricing decision support system for managing the request for quotation process under e-commerce business environment. *International Journal of Production Research*, 57(20), 6528–6551. doi:10.1080/00207543.2019.1566674
- Li, C., Chu, M., Zhou, C., & Zhao, L. (2020). Two-period discount pricing strategies for an e-commerce platform with strategic consumers. *Computers & Industrial Engineering*, 147, 106640. doi:10.1016/j.cie.2020.106640
- Liu, Z., Guo, H., Zhao, Y., Hu, B., Ji, X., Lang, L., & Shi, L. (2022). Optimal pricing decision of composite service offered by network providers in E-commerce environment. *Electronic Commerce Research*, 22(1), 177–193. doi:10.1007/s10660-021-09487-8
- Lv, J., Wang, Z., Huang, Y., Wang, T., & Wang, Y. (2020). How can e-commerce businesses implement discount strategies through social media? *Sustainability*, 12(18), 7459. doi:10.3390/su12187459

- Moerth-Teo, J. Y., Bobek, V., Horvat, T., & Milost, F. (2021). The Effects of Consumers' Buying Behavior on E-Commerce Pricing in the Highly Developed Emerging Market: The Case of Singapore. *International Journal of Economics and Finance*, 13(12), 1–55. doi:10.5539/ijef.v13n12p55
- Nemati, S., Rohani, R., Basiri, M. E., Abdar, M., Yen, N. Y., & Makarenkov, V. (2019). A hybrid latent space data fusion method for multimodal emotion recognition. *IEEE Access: Practical Innovations, Open Solutions*, 7, 172948–172964. doi:10.1109/ACCESS.2019.2955637
- Nguyen, D., Nguyen, K., Sridharan, S., Dean, D., & Fookes, C. (2018). Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition. *Computer Vision and Image Understanding*, 174, 33–42. doi:10.1016/j.cviu.2018.06.005
- Siddiqui, M. F. H., Dhakal, P., Yang, X., & Javaid, A. Y. (2022). A Survey on Databases for Multimodal Emotion Recognition and an Introduction to the VIRI (Visible and InfraRed Image) Database. *Multimodal Technologies and Interaction*, 6(6), 47. doi:10.3390/mti6060047
- Siriwardhana, S., Kaluarachchi, T., Billingham, M., & Nanayakkara, S. (2020). Multimodal emotion recognition with transformer-based self supervised feature fusion. *IEEE Access: Practical Innovations, Open Solutions*, 8, 176274–176285. doi:10.1109/ACCESS.2020.3026823
- Sullivan, Y. W., & Kim, D. J. (2018). Assessing the effects of consumers' product evaluations and trust on repurchase intention in e-commerce environments. *International Journal of Information Management*, 39, 199–219. doi:10.1016/j.ijinfomgt.2017.12.008
- Tong, W., Chen, W., Han, W., Li, X., & Wang, L. (2020). Channel-attention-based DenseNet network for remote sensing image scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4121–4132. doi:10.1109/JSTARS.2020.3009352
- Tseng, K. K., Lin, R. F. Y., Zhou, H., Kurniajaya, K. J., & Li, Q. (2018). Price prediction of e-commerce products through Internet sentiment analysis. *Electronic Commerce Research*, 18(1), 65–88. doi:10.1007/s10660-017-9272-9
- Wang, C. N., Dang, T. T., Nguyen, N. A. T., & Le, T.-T.-H. (2020). Supporting better decision-making: A combined grey model and data envelopment analysis for efficiency evaluation in e-commerce marketplaces. *Sustainability*, 12(24), 10385. doi:10.3390/su122410385
- Wu, C. H., Yan, Z., & Tsai, S. B. (2020). An empirical study on sales performance effect and pricing strategy for E-commerce: from the perspective of mobile information. *Mobile Information Systems*.
- Wu, X., Zheng, W. L., Li, Z., & Lu, B.-L. (2022). Investigating EEG-based functional connectivity patterns for multimodal emotion recognition. *Journal of Neural Engineering*, 19(1), 016012. doi:10.1088/1741-2552/ac49a7 PMID:35094982
- Xie, H., Ma, R. T., & Lui, J. C. (2018). Enhancing reputation via price discounts in e-commerce systems: A data-driven approach. *ACM Transactions on Knowledge Discovery from Data*, 12(3), 1–29. doi:10.1145/3154417
- Xu, W., Zhou, S., Zhao, K., Shi, Y., & Bingzhen, S. (2021). Optimal pricing model for non-instantaneous deterioration items with price and freshness sensitive demand under the e-commerce environment in China. *Kybernetes*, 51(2), 623–640. doi:10.1108/K-05-2020-0261
- Zhang, H. (2020). Expression-EEG based collaborative multimodal emotion recognition using deep autoencoder. *IEEE Access: Practical Innovations, Open Solutions*, 8, 164130–164143. doi:10.1109/ACCESS.2020.3021994

Ziqi Zhong is a PhD Candidate in Management - Marketing, fully funded by the LSE and the Lee Family PhD Scholarship. His research interest lies in digital marketing, mainly focusing on consumer interaction and sustainability innovation, with the aim to shed light on social welfare. Methodologically, he utilises different approaches, including empirical modelling, applied AI and behavioural experiments.