ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Characteristics of cross-border e-commerce marketing talent demand for agricultural products and the supply-side reform path of farmer training

Guangshi Pan[1,*] *and Mei Guo*[1]

[1] College of Business Administration, Tongling University, Tongling, Anhui, 244061, China

SUMMARY: *Farmers play a key role in the cross-border e-commerce sale of agricultural goods and the level of their professional literacy and the practicality of their functioning have a direct influence on end marketing results. With the competency demands of cross-border e-commerce marketing personnel as an analytical starting point, this paper gathers the data on the demand of talents in the online recruitment websites and extracts the corresponding texts using web-crawling tools. On such grounds, the LDA topic model is used to derive the main terms of e-commerce marketing talent as well as to group the most significant themes and the SOFM algorithm is additionally used to group the characteristics of the talent demand in order to discover the critical qualities necessary in the cross-border e-commerce marketing practitioners. The sample includes the recruitment ads of cross-border e-commerce marketing jobs found on mainstream Internet employment sites. Based on job duties and role characteristics, the agricultural product cross-border e-commerce positions are divided into four kinds, which include planning, promotion, research, and operation. Considering the demand trends and skills requirements in those categories, it is assumed in the paper that supply-side reforms in the training of farmers to agricultural-product cross-border e-commerce marketing need to be aligned with the characteristics of the product and the requirements of markets, as well as improve the quality of courses and extend training resources simultaneously.*

KEYWORDS: *cross-border e-commerce; marketing talent demand; SOFM algorithm; farmer training; supply-side reforms*

1 Introduction

Amid the continued advancement of China's agricultural supply-side structural reform, more participants in the industry have come to recognize the necessity of bringing high-quality agricultural products into broader overseas markets. Under this background, cross-border e-commerce has become a key bridge linking premium domestic production areas with global buyers [1-4].Compared with the traditional model of bulk trade, this approach does more to present product characteristics and cultural value, while better meeting the increasingly diverse and personalized demands of consumers worldwide[5, 6].

However, behind the rapid growth of the cross-border e-commerce sector, a continuing and hard-to-resolve problem remains the shortage of interdisciplinary professionals required by this field [7]. Effective marketing of agricultural products through cross-border e-commerce depends on versatile personnel who possess international trade competence as well as a clear understanding of cross-border e-commerce operations, its professional skills

*zzgdvicn@126.com
https://doi.org/10.65102/is2026400 Published: 30 April, 2026

requirements are even higher [8, 9]. Such roles require innovative interdisciplinary personnel who can track development trends in the agricultural products industry, understand product preferences and consumption habits in target overseas markets, and integrate e-commerce technology with international trade expertise[10].Although colleges and universities have recognized the urgency and importance of cultivating cross-border e-commerce talent, adjustments in their training systems still tend to lag behind actual industry needs. School-enterprise cooperation and practice-based zero distance cultivation should be implemented in a tangible way to cultivate highly skilled composite professionals that can fit into the requirements of industries and enterprises.

Yet another significant challenge that impacts cross-border e-commerce of agricultural produce is that there is a low quality level in several farmers and the existence of old-fashioned beliefs. To realize sustainable development in this industry, there is no alternative to adopting a supply-side reform pathway with a focus on efficient training of the farmers [11-13]. This will require government agencies to enhance the talent-training mechanism related to the new rural e-commerce, reinforce the guidance role of farmer education, enhance collaboration with companies, combine quality resources, develop platforms to train farmers and create a multi-level training framework [14-16]. Moreover, entrepreneurial projects could provide farmers with real-life business practices opportunities to gain experience in the field of e-commerce, learn about the industry, and develop a sense of professionalism in the process of practical experience. [17, 18].

The recruitment data on cross-border e-commerce marketing positions were initially gathered using the web-crawling technology in this research. The LDA topic model was subsequently used to identify the underlying attributes and central attributes of the job-related texts. On that basis, the SOFM algorithm was developed to cluster thematic information about talent characteristics and hence develop a combined text-mining and clustering system to examine the demand attributes of cross-border e-commerce marketing personnel. The analysis of clustering outcomes revealed how many major themes of talent-characteristics could be detected. It also investigates the main aspects of starting salaries as perceived by talent-demand and the specific characteristics of various positions as seen by recruitment-demand. Combining the analytical results and the talent portrait of cross-border e-commerce marketing personnel provides a final way forward to supply-side reforms in the training of farmers.

2 Text mining and clustering of e-commerce marketing talent demand characteristics

2.1 Text mining key technology

2.1.1 Web crawling techniques

E-commerce marketing talent job data collection is realized using web crawler technology, the essence of web crawler is a computer program that obtains web information from web pages according to certain rules. The information on the Internet is always in a state of dynamic updating, the web crawler makes the work of crawling simple, people only need to clarify their crawling target, formulate good crawling rules and storage methods. According to the complexity of the Web page to choose the appropriate crawler technology, in general the Web crawler is divided into three categories: general-purpose Web crawler, thematic Web crawler, deep crawler. The data collection process of e-commerce marketing talent positions based on web crawler is as follows: first, select the target recruitment website for crawling, and capture the URL of the target recruitment webpage to form a queue; second, set the crawling rules (crawling depth and width); third, capture the target recruitment webpage and analyze the content and characteristics of the talent information, and at the same time, continuously capture the URL of the new target webpage and add it to the queue until all the data are acquired. In order to ensure that the crawler can accurately traverse the webpage information it needs to obtain, it sets up a part of important URLs in advance, and traverses these URLs under the already determined crawling rules. The general flow of the web crawler is shown in Figure 1.

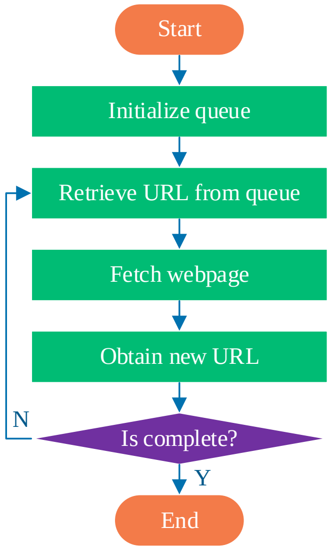

Figure 1: The traditional web crawler process

2.1.2 TF-IDF

The TF-IDF model is a statistical way to measure the importance of a word or phrase in a document. It is usually considered that the same word or phrase occurs relatively more frequently in longer documents than in shorter ones, in order to solve this problem, the word frequency is normalized.The TF value is calculated as expressed in equation (1):

$$TF = \frac{count(w,d)}{size(d)} \tag{1}$$

d denotes the document, w denotes the word or words, $size(d)$ denotes the total number of words, and $count(w,d)$ denotes the number of times the word appears in the document.

The IDF expression is shown in equation (2):

$$IDF(w,D) = \log\left\{\frac{N - \text{Total Document}}{1 + count\{w\_in\_d_i \mid i = 1,2,3,\cdots,N\}}\right\} \tag{2}$$

The use of plus 1 in the denominator is a common way to prevent the denominator from being 0. The more times a word occurs the larger the denominator will be and the more the IDF will converge to 0. The TF-IDF is expressed as in equation (3):

$$TF - IDF = TF * IDF \tag{3}$$

As shown in Equation (3), if a certain word appears in all documents with higher frequency words, then TF-IDF will be larger, and meaningless high-frequency words are filtered by judging the value of TF-IDF. Similarly, the value of TF-IDF can be judged to determine whether a talent information feature word is a keyword.

2.1.3 Mutual information and left-right entropy

This paper uses left-right entropy discrimination and mutual information to perform keyword identification in job-talent data. Mutual information is defined as equation (4):

$$PMI(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{4}$$

where $p(x,y)$ denotes the probability that two words co-occur and $p(x)$ denotes the probability that a single word appears. Mutual information reflects the strength of association between two words.

The left and right entropy measures are used to assess the contextual freedom of the candidate words. To any given pair of words, the entropy values on either side are computed and a higher entropy value will suggest more emphatically that a new word has been created. As entropy describes uncertainty, it can be said that a higher value of entropy denotes a higher level of uncertainty and thus points to the higher amount of left-right collocation information of the word pair. It is defined in Eq. (5):

$$E_L(W) = -\sum_{\forall \alpha \in A} P(\alpha W) \log_2 P(\alpha W | W) \tag{5}$$

2.1.4 Text pattern mining

The text pattern analysis consists of methods of text clustering, text classification, correlation analysis, and association analysis. Text clustering is an example of unsupervised learning that clusters documents into a number of classes based on textual data and relationships between internal features, without any prior definition of the classes to be formed. Popular algorithms used are K-means and EM.Text classification, however, is a common form of supervised learning. It builds classification rules using previous studies and subsequently classifies texts into similar categories through the discovery of their characteristic features. Text categorization is commonly performed by algorithms, such as SVM, KNN, and AdaBoost. Correlation analysis is applied to assess the level of similarity between high-frequency words and respective documents. The correlation value exceeding one is indicative that these kinds of words should be examined more closely. Also, a larger correlation value means that the difference between the two types of words is more significant, which implies that the given word or the word type is highly required by the position in question rather than in other positions. Thus, text correlation analysis assists in discovering the connections between high frequency words in a corpus and uncovering meaningful association trends out of them.

2.2 LDA Subject Modeling

A document may be created within the LDA framework by initially sampling a topic through a probability distribution and subsequently choosing a word based on the associated probability. It can be done using repeated iterations of this process until a full-fledged article is formed.

Suppose that there are m documents after lexical preprocessing, as illustrated in Fig. 2.

Here, w_{ij} denotes the j th word in the i th document, and z_k represents the k th topic. The probability of generating the j th word in document D_m is given in Eq. (6), where $P\left(w_{mj}|z_k\right)$ and $P\left(z_k|D_m\right)$ refer to the topic-word probability and the document-topic probability, respectively.

$$P\left(w_{mj}|D_m\right)=\sum_{k=1}^{K}P\left(w_{mj}|z_k\right)P\left(z_k|D_m\right) \tag{6}$$

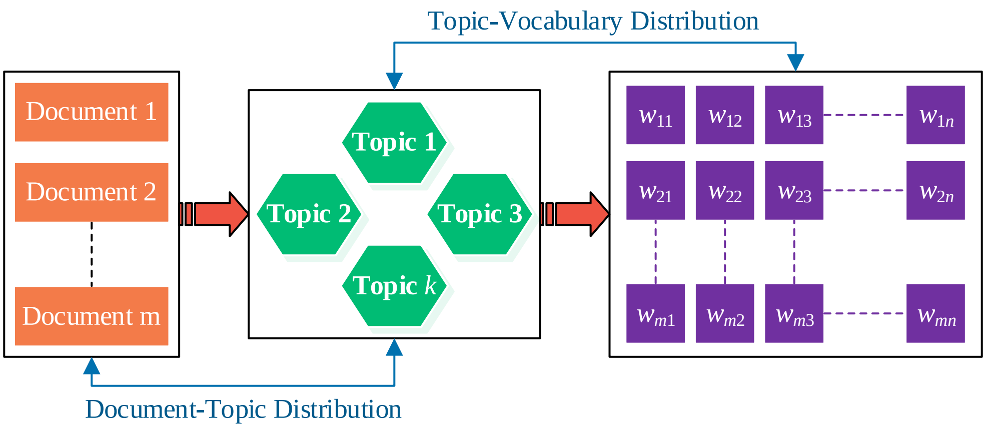

Figure 2: The PLSA model concept

Following this idea, a topic model can infer the topic distribution of an article from the distribution of words. The LDA model is conceptually close to PLSA, but differs from it in that *Dirichlet* priors are introduced for the document-topic and topic-word distributions. Because the *Dirichlet* distribution can be regarded as a high-dimensional extension of the *Beta* distribution, its derivation process is not discussed here in detail; only the corresponding probability distribution function is presented in Eq. (7)

$$Dir\left(\vec{p}|\vec{a}\right)=\frac{1}{B(\vec{a})}=\frac{\prod_{k=1}^{K}Ga(a_k)}{\tilde{O}_{k=1}^{K}Ga(a_k)}\prod_{k=1}^{K}p_k^{a_k-1} \tag{7}$$

where $\vec{a}=\left(a_1,a_2,\cdots,a_k\right)$ is the parameter of the *Dirichlet* distribution and $a_1,a_2,\cdots,a_k>0$; $B(\vec{a})$ denotes the normalization constant of the *Dirichlet* distribution. According to the derivation idea of the model, the topic under the document and the distribution of words under the topic can be obtained from the prior distribution *Dirichlet* in turn, so as to obtain the probability value of the corresponding topic of the document according to the probability of the corresponding topic of the word, and finally classify or cluster the topic according to the topic characteristics to complete the classification and clustering of the text of the document, which can be solved by the *Gibbs* sampling algorithm, the variational extrapolation *EM* algorithm for the model.

The specific steps are shown below:

(1) Select a document D_i according to the prior probability q_d;

(2) Generate the distribution of the topic of document D_i from the hyperparameter for

a topic *Dirichlet* distribution q_d;

(3) Generate the topic $Z_{d,n}$ of the jth word of document D_i from a polynomial distribution q_i of topics;

(4) Generate the distribution $W_{d,n}$ of words corresponding to topic $Z_{i,j}$ from the *Dirichlet* distribution b_k with hyperparameter h;

(5) Sampling the final generated words from the polynomial distribution of words.

2.3 SOFM algorithm

2.3.1 Basic Ideas

Self-organizing feature map neural network (SOFM) is an unsupervised clustering network that relies on competitive learning. Its main principle is to find out the underlying regularities and important characteristics of the sample data and achieve self-organization with adaptive optimization of network parameters and topology. The functioning of SOFM algorithm can be roughly divided into two steps, which are learning and work. In learning step, training examples are randomly fed into the network. At any input pattern, one neuron of the output layer responds the most strongly and is the winner neuron, but during the first step of learning the location of the maximally responsive neuron has not been fixed in the output layer yet. With the change in the category of the input pattern, the location of the winner neuron in the two-dimensional plane shifts. The neighboring neurons of the winner also experience lateral inhibition, so the weight vectors associated with the winner neuron and the neurons in its superior neighborhood are modified differently towards the input pattern. The adjustment becomes stronger as the neuron gets closer to the winner. Through repetitive training on large numbers of samples, self-organization occurs as the connection weights are adaptively changed and neurons in the output layer will eventually be sensitive to certain input patterns, and the associated connection weights will develop into the center vectors of these patterns. When two classes of input patterns have similar features, the neurons that represent them are also placed close to each other. This leads to an orderly feature map being created in the output layer, which indicates the distribution of sample pattern classes.

The basic procedure of the SOFM algorithm is as follows:

(1) Initialize the connection weights w, learning rate η_0, and neighborhood radius δ_0.

(2) Sample the input data and perform steps (3)–(6) for each sample.

(3) Determine the winning neuron according to $i(X) = \arg\min_j \left\| X - W_j \right\|, j = 1, \cdots k$.

(4) Update, according to Eq. (8), the connection weights of the winning neuron and all neurons within its neighborhood, whereas neurons outside the neighborhood radius remain unchanged.

$$W_j(n+1) = \begin{cases} W_j(n) + \eta(n) h_{j,i(X)}(n)(X - W_j) \\ W_j(n) \end{cases} \tag{8}$$

(5) Adjust parameters such as learning rate and neighborhood radius.

(6) Restart from (2) until the SOFM algorithm converges or reaches the maximum number of iterations.

2.3.2 Mathematical description

The SOFM learning process has three phases, which are competition, cooperation, and adaptation.

(1) During the competition stage, the network determines the winning neuron corresponding to a given input pattern. For each sample, all neurons evaluate the value of their own discriminant functions. Suppose the dimension of the input vector space is m. Then the input vector can be written as: $X=\left[x_1,x_2,\cdots,x_m\right]^T$, and the weight vector of neuron j can be expressed as: $W_j=\left[w_{j1},w_{j2},\cdots,w_{jm}\right]^T$, $j=1,2,\cdots,k$. where k denotes the total number of neuron nodes in the output layer of the SOFM network. The response of neuron j can be described in two forms.

1) The inner product form is shown in equation (9):

$$\left\langle W_j,X\right\rangle=\sum_{j=1}^{k}w_{ji}x_i=W_j^TX \tag{9}$$

2) The Euclidean distance form is shown in equation (10):

$$d_j=\left\|X-W_j\right\|=\left(\sum_i\left(x_i-w_{ji}\right)^2\right)^{1/2} \tag{10}$$

These two functions are used to measure the similarity between the input vector X and the weight vector W_j. A larger inner product or a smaller distance d_j indicates a higher degree of similarity between X and W_j. It has been shown that, once the input vector and the weight vector are normalized, maximizing the inner product is equivalent to minimizing the Euclidean distance between them.

When $\left\|W_j\right\|=1$, $\left\|X\right\|=1$, there is equation (11):

$$\begin{aligned}d_j^2&=\left\|X-W_j\right\|^2=\left\langle X-W_j,X-W_j\right\rangle=\left\langle X,X\right\rangle-2\left\langle X,W_j\right\rangle+\left\langle W_j,W_j\right\rangle\\&=\left\|X\right\|^2-2\left\langle X,W_j\right\rangle+\left\|W_j\right\|^2=2-2\left\langle X,W_j\right\rangle\end{aligned} \tag{11}$$

Thus, there is equation (12):

$$i\left(X\right)=\arg\min_j\left\|X-W_j\right\|,\ j=1,\cdots k \tag{12}$$

According to Eq. (12), for neuron recognition i, $i\left(X\right)$ is taken as the target under consideration. The neuron i that satisfies this condition is referred to as the best-matching unit, namely the winning neuron.

(2) In the cooperation stage, the excitation center of the winning neuron is determined, and the topological neighborhood of the winning neuron node is then calculated. Let $h_{j,i}$ denote the topological neighborhood centered on winning neuron node i, and let $d_{j,i}$ represent the lateral distance between winning neuron node i and excitatory neuron node j. It is assumed that the topological neighborhood $h_{j,i}$ is a single-peaked function of the lateral

distance $d_{j,i}$, and that it meets the following conditions.

1)The topological neighborhood $h_{j,i}$ is symmetric about the maximum point defined by $d_{j,i} = 0$, that is, the maximum value is attained at winning neuron i when $d_{j,i}$.

2)The magnitude of the topological neighborhood $h_{j,i}$ decreases monotonically as the lateral distance $d_{j,i}$ increases, and approaches zero when $d_{j,i} \to \infty$, which is a necessary condition for convergence.

A common form of $h_{j,i}$ satisfying these conditions is the Gaussian function, as shown in Eq. (13).

$$h_{j,i(X)} = \exp\left(-\frac{d_{j,i}^2}{2\delta^2}\right) \tag{13}$$

It is independent of the position of the winning neuron and therefore remains invariant under translation. Figure 3 presents the Gaussian neighborhood function, in which δ denotes the effective width of the topological neighborhood and reflects the degree to which excitatory neurons near the winning neuron participate in the learning process.

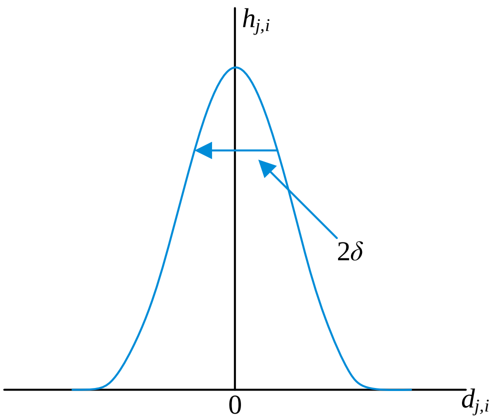

Figure 3: Gaussian neighborhood function

In Eq. (13), δ is the neighborhood radius, δ is dynamically changing, and its change rule is shown in Eq. (14):

$$\delta(n) = \delta_0 \exp\left(-\frac{n}{\tau_1}\right), n = 0, 1, 2 \cdots \tag{14}$$

where δ_0 denotes the initial neighborhood radius and τ_1 denotes that time is a constant. Substituting (14) into (13) there is equation (15):

$$h_{j,i(X)}(n) = \exp\left(-\frac{d_{j,i}^2}{2\delta^2(n)}\right), n = 0, 1, 2 \cdots \tag{15}$$

$h_{j,i(X)}(n)$ is the neighborhood function.

(3) The adaptive stage uses an altered Hebbian learning rule that updates the weight

vectors of the winning neuron and the neurons that are within its topological neighborhood. By this adaptive process, excitatory neurons can correctly adapt their synaptic weights such that the value of the discriminant function associated with the present input pattern is maximized. In turn, the winning neuron response to subsequent input patterns with similar features is reinforced even more.

With repeated training on the sample data, neighborhood-based updating causes the synaptic weights to gradually conform to the distribution of the input vectors. In this formulation, the Hebbian assumption is revised by introducing a forgetting term $g(y_i)w_j$, where $g(y_i)$ is a positive scalar function of the response y_i. The only requirement imposed on $g(y_i)$ is that the constant term in its Taylor series expansion is equal to zero, as shown in Eq. (16).

$$g(y_i) = 0, y_i = 0 \tag{16}$$

Given such a function, the change in the neuron j weight vector in the grid is expressed as equation (17):

$$\Delta W_j = \eta y_j x - g(y_j)W_j \tag{17}$$

where η is the learning rate parameter of the algorithm, $\eta y_j x$ is the Hebb term, and $g(y_j)w_j$ is the forgetting term. In order to satisfy Eq. (16), a linear function is chosen for $g(y_j)$ as in Eq. (18):

$$g(y_j) = \eta y_j \tag{18}$$

Eq. (17) can be further simplified by setting up Eq. (19):

$$y_j = h_{j,i(X)} \tag{19}$$

Substituting Eq. (18) and Eq. (19) into Eq. (17) yields Eq. (20):

$$\Delta W_j = \eta h_{j,i(X)} \left(X - W_j \right) \tag{20}$$

Assuming that the weight vector of neuron j at moment n is $w_j(n)$, the update weight vector $w_j(n+1)$ at moment $n+1$ is defined as equation (21):

$$W_j(n+1) = W_j(n) + \eta(n) h_{j,i(X)}(n) \left(X - W_j \right) \tag{21}$$

where $\eta(n) = \eta_0 \exp\left(-\dfrac{n}{\tau_2} \right)$ is the learning rate, η_0 is the initial learning rate, and τ_2 is a constant indicating time.

3 Supply-side reform of farmer training that integrates the needs of marketing personnel

3.1 Processing of talent data information for cross-border e-commerce marketing positions

3.1.1 Data collection results

Internet/electronic job postings of e-commerce marketing professionals were collected nationwide, and a total of nearly 200,000 data were acquired. By screening and reclassifying the data, finally 168,401 data were retained for analysis.

The collected data fields include the following 12: job position, company name, company nature, company size, monthly salary of the position, posting date, work experience, number of recruits, work location, educational requirements, job responsibilities, job requirements.

The collected data are stored locally in the form of Excel tables, called table headers by field names, in which, except for job duties and job requirements, which are paragraph texts, the remaining fields are mainly in the form of numerical values and phrases, most of which have specific value ranges and are relatively structured data. Therefore, the following analysis will focus on the unstructured text data, especially the job requirement text.

3.1.2 Determination of the number of topics

The LDA topic model is used to construct models and extract topics for the job duties and job requirements in the cross-border e-commerce marketing job information, so as to obtain the keywords corresponding to each topic. When the number of topics k is 4, the clustering of the extracted topic keywords is shown in Fig. 4. The circles with black edges in the figure are the keywords under different categories, and the circles without black edges of the same color are the categories to which they belong. The distance between the circles of different colors without black edges represents the similarity of the two themes, the shorter the distance, the higher the similarity. Among them, the central keyword of theme 1 is product, market, the central keyword of theme 2 is media, publicity, the central keyword of theme 3 is mining, analysis, and the central keyword of theme 4 is customer, sales. And under 4 themes, the distance between keyword clusters is far and there is no crossover phenomenon. It shows that when the number of themes is 4, the similarity between themes is weak and the classification effect is better.

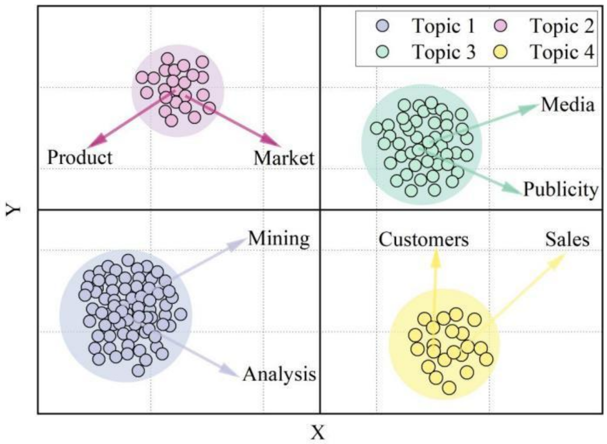

Figure 4: Visualization of LD theme Model Effect (k=4)

When the number of topics k is 5, the clustering of the extracted topic keywords is shown in Figure 5. Among them, topic 3 and topic 5 have obvious cross overlap phenomenon, and the distance between the two and topic 4 is relatively close, indicating that the repetition rate of the keywords under the category of topic 3 and topic 5 (whose center keywords are data and brand) is high, and they are more similar to the topic 4, and the effect of the topic extraction is weaker. Therefore, this paper determines that the number of themes k of cross-border e-commerce marketing talent job characteristics is 4 for LDA theme extraction.

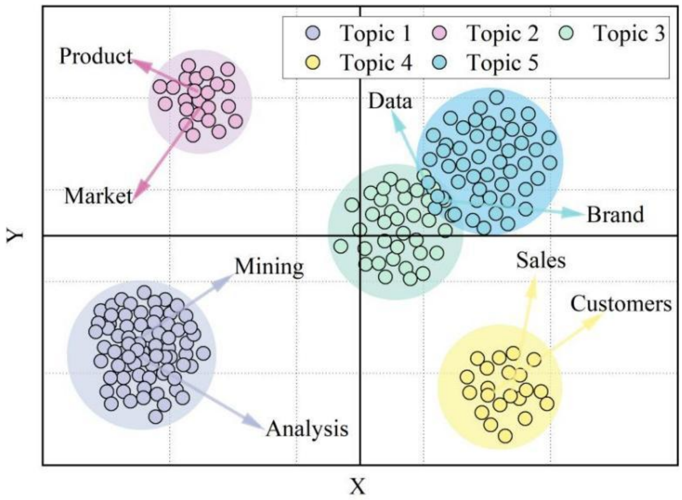

Figure 5: Visualization of LD theme Model Effect (k=5)

3.2 Characterization of talent demand for cross-border e-commerce marketing positions

3.2.1 Characterization of the impact of the starting salary of post talent

Combining the actual situation with the existing research, 30 more important parameters influencing the starting salary characteristics of jobs were selected: 0=first-tier cities, 1=second-tier cities, 2=other cities, 3=new first-tier cities, 4=publicly listed companies, 5=startups, 6=state-owned enterprises, 7=foreign enterprises, 8=private firms, 9=non-profit organizations, 10=1000-5000 people, 11=more than 10,000 people, and 12=150-500 people, 13=500-1000 people, 14=less than 150 people, 15=More than 10 years of experience, 16=1 year of experience, 17=Master's Degree, 18=3-4 years of experience, 19=5-7 years of experience, 20=8-9 years of experience, 21=Current Freshers, 22=No experience required, 23=Secondary, 24=Other, 25=Doctorate, 26= College, 27=Bachelor's degree, 28=2 years of experience.

The importance ranking of the impact characteristics of cross-border e-commerce marketing positions is shown in Figure 6. Among them, the feature "0= first-tier city" and "21= current student or fresh graduate" have an impact F value of over 150 on the starting salary, while the feature "22= no experience required", "15= over 10 years of experience" and "17=2 years of experience" rank 3rd to 5th respectively (F>120). It indicates that the work location, work experience and educational background are important influencing factors for the starting salary of cross-border e-commerce marketing talents.

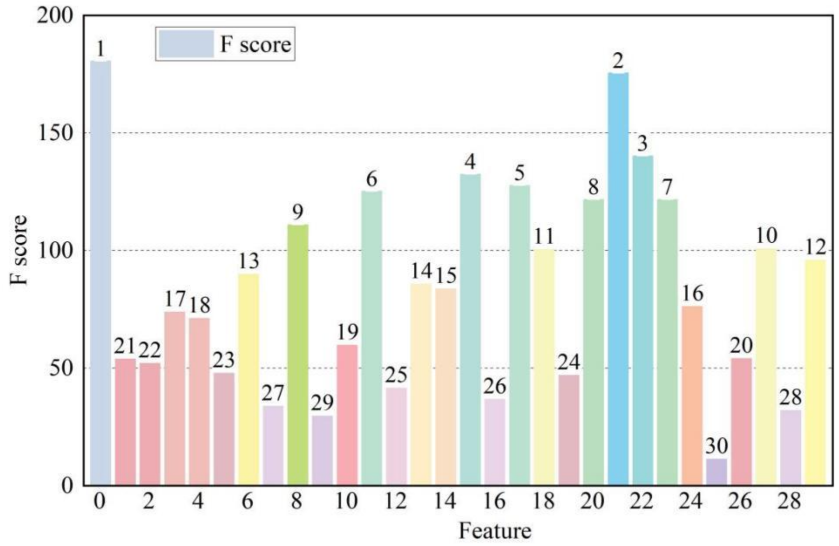

Figure 6: Ranking of job importance characteristics

3.2.2 Characteristics of Job Talent Demand Based on SOFM Clustering Algorithm

The clustering effect of the three clustering algorithms is shown in Fig. 7. Overall, the contour coefficients of the three algorithms show a fluctuating downward trend with the increase of the number of clusters, and reach the highest point when the number of clusters is three. The highest point is reached when the number of clusters is 3. In terms of specific numerical performance, the SOFM algorithm is the best (0.15~0.72), the bag-of-words model integrating TF-IDF is the second best (0.14~0.65), and the Word2ec model is the last (0.09~0.46), which not only verifies the analytical result that the optimal number of clustered topics should be 4, but also shows the superiority of the clustering algorithm of this paper.

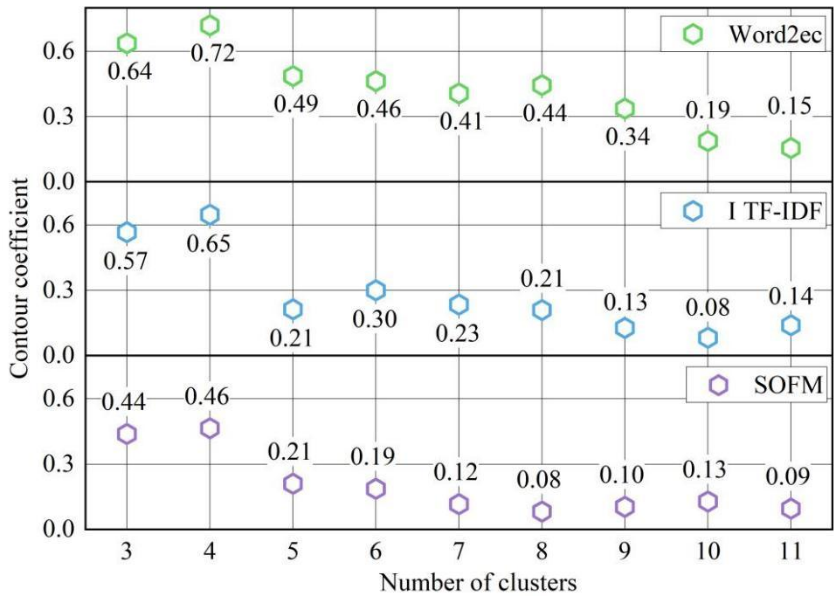

Figure 7: Comparison of clustering algorithm results

To determine the names of each job type, the top 10 keywords of the job name for each category were extracted and the word frequency was statistically analyzed. The results are shown in Table 1. It can be seen that e-commerce marketing positions are subdivided into four categories: planning, promotion, research and operation. Among the planning positions, the three texts of product project initiation (586), goal setting (536), and product planning (471) have a relatively high frequency of appearance. Among the promotion positions, the texts of product promotion (428), digital marketing (391), and brand planning (344) rank the top three in terms of frequency of appearance Among the research positions, the text frequency of demand insight (305), data mining (279), and data analysis (245) is relatively high. Among the operation positions, the text frequency of business processing (504), product sales (461), and customer maintenance (405) is relatively high. The text content with a relatively high frequency of appearance for different types of positions indicates that the skill requirements for talents in cross-border e-commerce marketing positions are mainly concentrated in three aspects: product design, product promotion and customer service.

Table 1: Key words of job category

Category 1		Category 2	
Planning	Word frequency	Promotion	Word frequency
Product Launch	586	Product Promotion	428
Goal Setting	536	Digital Marketing	391
Product Planning	471	Product Publicity	344
Market Development	360	Media Resources	263
Marketing Strategy	354	Product Story	258
Product Establishment	347	Sales	253
Product Design	295	Brand Building	215
Product Services	246	Public Relations	180
Product Management	102	Events	74
Resource Allocation	81	Brand Planning	59
Category 3		Category 4	
Pesearch	Word frequency	Operation	Word frequency
Demand Insight	305	Business Processing	504
Data Mining	279	Product Sales	461
Data Analysis	245	Customer Maintenance	405
Assistant	187	Product Services	310
Algorithm	184	Team Building	304
Specialist	180	Team Management	298
Customer Needs	153	Communication	254
Researcher	128	Team Coordination	212
Consult	53	Platform Maintenance	88
Market Tracking	42	Customer Follow-up	70

3.2.3 Job skill analysis based on talent characteristic clustering

Visualized cross-border e-commerce marketing job talent demand characteristics are shown in Figure 8, which is consistent with the analysis above, and its job talent demand characteristics are divided into 4 categories. The boundaries between clusters and clusters are clearly delineated, there is no keyword crossover phenomenon, and the clustering effect is better.

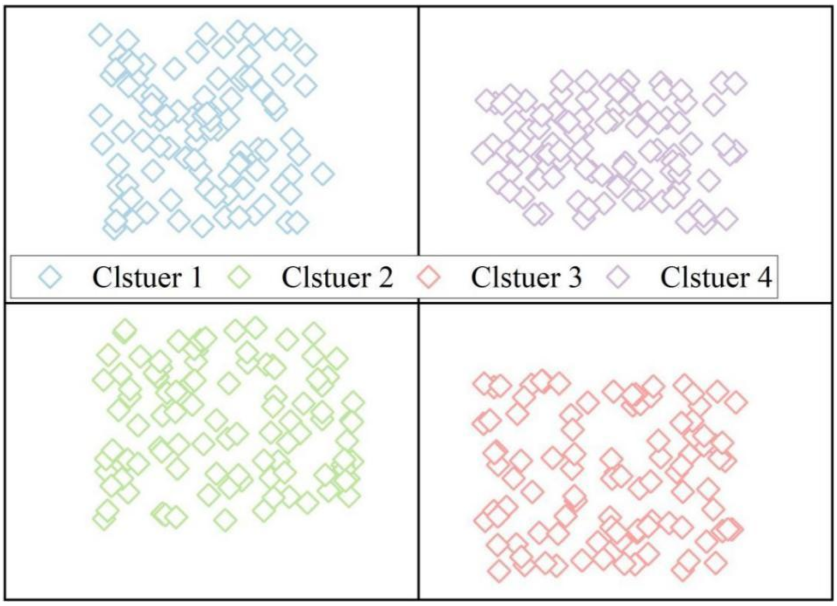

Figure 8: Characteristics of talent demand for e-commerce marketing positions

Drawing four categories under the cross-border e-commerce marketing job skills demand word cloud situation is shown in Figure 9, from the word cloud of each category can be seen in the category under the position under the core skills that talents should have. Combined with the analysis in the previous subsection, agricultural cross-border e-commerce marketing planning talents should have keen market demand insight ability, good market strategy deployment ability and product planning ability, promotion talents need to have solid media promotion ability and excellent channel development ability; research talents are biased towards data mining and analysis skills; operation talents need to have good language communication and team coordination ability, to support their customer development, maintenance and platform operation. The operation talents need to have good language communication and team coordination skills to support their customer development, maintenance and platform operation work.

Figure 9: Clustering result word cloud

3.3 Supply-side reform of farmers' training under the characteristics of talent demand

In today’s cross-border e-commerce sector for agricultural products, the core difficulty is how to build an effective connection between farmers and online trade platforms. Focusing on the supply side of skills development, this paper discusses ways to strengthen farmers’ professional competence and digital literacy in agricultural sales. The objective is to ease the structural mismatch on the training supply side by increasing both the amount and the standard of training provision, while establishing a practical and well-designed training framework for farmers, thereby improving their marketing capability and supporting the steady expansion of cross-border online trade in agricultural products.

Training materials should not remain confined to the narrow topics of “agricultural products” and “e-commerce marketing.” Instead, they need to include a broader range of e-commerce-related issues, such as consumer psychology and poster design. In this way, farmers can gradually develop a deeper understanding of digital trade platforms and customer demand, while continuously strengthening their marketing competence in this field.

The quality of training content should match market demand, product features, and platform operating mechanisms as closely as possible. By examining user characteristics on online commerce platforms and studying market demand for agricultural products, it becomes possible to design high-quality training content for the cross-border marketing of farm products. At the same time, it is necessary to follow farmers’ learning progress and coursework performance, formulate short-term and long-term teaching plans, and implement differentiated instruction. This approach can improve practical teaching materials and systematic teaching methods, thus raising both the efficiency and the overall quality of farmer training.

4 Conclusion

The proposed supply-side reform pathway is meant to address the lack of specialization among farmers in cross-border e-commerce marketing of agricultural products through talent portraits based on cross-border e-commerce marketing positions that will serve as guidance in capability training. The LDA-based topic modelling divides related occupations into four groups, including planning, promotion, research, and operation. When the influential factors were analyzed, it was found that the three main factors that determine the starting salary are workplace, work experience, and educational background (F>120). In various topic clusters, the SOFM algorithm yields coefficients of profiles between 0.15 and 0.72, allowing clustering of demand features in the marketing posts and determining the essential competencies needed by various types of jobs. Based on these, the training of farmers must be in line with the demands of the market, offer additional themes related to cross-border e-commerce marketing and enhance the ability of the farmers professionally through the enhancement of the quality and the magnitude of training offered.

Funding

This work was supported by Key Project of Philosophy and Social Science Research in Anhui Province's Universities in 2023: A Study on the Path and Policy of Organic Connection between Anhui Small Farmers and Modern Agriculture: From the Perspective of Small Farmers as the Main Body (2023AH051651).

About the Author

Guangshi Pan, male, born in Hefei, Anhui. People's Republic of China, 1973. He obtained a master's degree from Yunnan University for Nationalities. Now, he is currently working at the School of Business Administration of Tongling University. His research interests include agriculture, rural areas, and human resource management.

Mei Guo, female, born in Ma'anshan, Anhui. People's Republic of China, 1982. She obtained a master's degree from Yunnan University for Nationalities. I am currently working at the School of Business Administration of Tongling University. Her research interests include agriculture, rural areas, and human resource management.

References

[1] Qi, X. (2024). Leveraging e-commerce for transnational sales-marketing of rural agricultural products: the impact of digital platforms and cross-border logistics. International Journal of eBusiness and eGovernment Studies, 16(2), 171-193.

[2] Wang, G., Zhang, Z., Li, S., & Shin, C. (2023). Research on the influencing factors of sustainable supply chain development of agri-food products based on cross-border live-streaming e-commerce in China. Foods, 12(17), 3323.

[3] Zhou, K., Lu, F., Ruan, Y., Jiang, X., & Yu, L. (2023). Research on the influence and mechanism of cross-border E-commerce on the quality of agricultural products exported by China. Journal of Global Information Management (JGIM), 31(6), 1-23.

[4] Yue-ming, Z., & Cheng, Z. (2020, August). The Impact of Cross-border E-commerce on China’s Agricultural Products Export:—An Empirical Study Based on Big Data Processing. In 2020 International Conference on Big Data and Social Sciences (ICBDSS) (pp. 106-110). IEEE.

[5] Song, Z. (2023). Patterns and Prospect of Cross-border E-commerce of Agricultural Products in China. Advances in Economics and Management Research, 6(1), 42-42.

[6] Jia, Y., & Yundendorj, K. (2024). Research on the High-Quality Development of Cross-border E-commerce for Agricultural Products in China and Mongolia under the Perspective of Digital Empowerment. Pakistan Journal of Life & Social Sciences, 22(2).

[7] Zhang, P. (2023). Current Status and Strategic Research on the Development of Agricultural Products in Guangdong Province under the Background of Cross-border E-commerce. Agricultural & Forestry Economics and Management, 6(3).

[8] Jun, Y., Xiao-Hui, S., & Yan, G. (2021). The “Competition” for Training Cross Border E-commerce Talents in China on the Background of “Internet Plus”. Journal of Higher Education Theory & Practice, 21(6).

[9] Lu, X., & Zhang, S. (2017, June). Analysis on the Strategy of Cross-border E-commerce Talents-Based on the analysis of policy content. In 2017 International Conference on Management Science and Management Innovation (MSMI 2017) (pp. 54-58). Atlantis Press.

[10] Zhang, X., & Zhang, W. (2024). Research on the High-Quality Development of Agricultural Product Cross-Border E-Commerce Under Rural Revitalization. Asia Pacific Economic and Management Review, 1(3).

[11] Zhang, J. (2024). A Study on Strategies for the Collaborative Development of Multi-Subject Cross-border E-commerce in Rural Areas. Academic Journal of Business & Management, 6(3), 243-249.

[12] Wang, G., Hou, Y., & Shin, C. (2023). Exploring Sustainable Development Pathways for Agri-Food Supply Chains Empowered by Cross-Border E-Commerce Platforms: A Hybrid Grounded Theory and DEMATEL-ISM-MICMAC Approach. Foods, 12(21), 3916.

[13] Li, X., Li, Y., & Chen, Z. (2024). Impact of rural E-Commerce participation on farmers' household development resilience: evidence from 1229 farmers in China. Agriculture, 14(5), 692.

[14] Chen, C. L., Ding, Y., Huang, T. F., & Huang, M. L. (2023). Research on cross-border e-commerce marketing strategy and entry mode of agricultural small-and medium-sized enterprises in Taiwan. International Journal of Agriculture Innovation, Technology and Globalisation, 3(4), 370-387.

[15] Chen, W. H., Lin, Y. C., Bag, A., & Chen, C. L. (2023). Influence factors of small and medium-sized enterprises and micro-enterprises in the cross-border e-commerce platforms. Journal of Theoretical and Applied Electronic Commerce Research, 18(1), 416-440.

[16] Fang, M., & Huy, D. T. N. (2023). Building a cross-border e-commerce talent training platform based on logistic regression model. The Journal of High Technology Management Research, 34(2), 100473.

[17] Feng, Y., & Koompai, S. (2022). Demand Analysis and Training Mode Research of E-Commerce Talents in Guangxi Cross Border E-Commerce Comprehensive Experimental Zone. Science, Technology, and Social Sciences Procedia, 2022(2), CiM23-CiM23.

[18] Nasri, N., & Iskandar, Y. H. P. (2021). Pedagogical Factors Influencing Training of E-Commerce Entrepreneurs: The Malaysian Case. TEM Journal, 10(3), 1274.