Journal of Electronic Research and Application, 2026, Volume 10, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Construction of an E-commerce Sales Prediction Model Based on Deep Learning

Changjun Lv

School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, Shandong, China

Abstract: Due to the characteristics of e-commerce sales data, high dimensionality, diverse types, and nonlinear relationships, traditional models exhibit insufficient adaptability. This study aims to construct a deep ensemble prediction model to provide a scientific support for enterprise inventory and marketing decisions. The model integrates convolutional neural networks (CNN) with an improved weighted deep forest (WDF). A multi-granularity scanning mechanism enhances perception of local spatiotemporal features; a binary adaptive differential evolution algorithm dynamically selects key features; and Bayesian optimization precisely tunes model hyperparameters. Experiments are conducted using large-scale data from mainstream e-commerce platforms. Results show that the model achieves a mean absolute percentage error (MAPE) of only 6.8% on the test set, outperforming baseline models such as ARIMA, random forest, and XGBoost in capturing nonlinear trends and long-term dependencies. This adaptive ensemble model effectively addresses feature redundancy and model mismatch issues, significantly improving the stability of sales prediction and providing reliable technical support for precise e-commerce operations.

Keywords: Deep learning; Random forest; Bayesian optimization; Adaptive algorithm

Online publication: Jun 29, 2026

1. Introduction

In the era of big data, the integration of Internet technology and e-commerce has become increasingly close. Hence, accurately predicting whether users will purchase products has become key to enhancing enterprises' market competitiveness, and ensemble learning has made significant progress in sales forecasting. When dealing with data characterized by high dimensionality, diverse types, and complex relationships, traditional models suffer from insufficient self-adjustment capability and high dependence on manual selection of key features. This study constructs an adaptive prediction model that combines evolutionary algorithms with deep ensemble learning, incorporating data from diverse sources and types (including user behavior, product attributes, and market trends). Drawing on model optimization experience from medical imaging, the approach aims to improve adaptability and stability in multi-type data scenarios [1,2]. The model employs a binary adaptive differential evolution algorithm for automatic selection of important features and uses metalearners to adjust the combination of base learners, thereby providing more scientific decision support for e-commerce enterprises.

2. Literature review

Sales prediction forms the foundation of supply chain management and inventory control. Research methods have evolved from traditional statistical models to deep learning architectures. Deep learning excels at autonomously mining nonlinear and complex patterns in data, demonstrating clear advantages in handling e-commerce data with instability and long-term variations [3]. Current approaches commonly use long short-term memory networks (LSTM) to capture dependencies in time series or convolutional neural networks (CNN) to identify local important patterns. Building on this foundation, new methods combine CNN with improved weighted deep forests (WDF). Sliding-window multi-angle scanning enhances perception of local features, while multi-level forest structures combined with cross-validation strategies alleviate performance limitations of single models in multi-feature scenarios.

3. Theoretical foundations

3.1. Basic principles of deep learning

Deep learning offers clear advantages in e-commerce sales prediction. Leveraging convolutional neural networks, models can autonomously extract spatiotemporal patterns from sales data, significantly reducing the burden of manual preprocessing. In the multi-granularity scanning framework, sliding windows divide raw features into subsets, effectively increasing the dimensionality and richness of information representation. Data are processed by cascaded forests composed of random forests and completely random forests, with weighted ensemble performed via k-fold cross-validation to produce final predictions [4]. This architecture integrates feature extraction and hierarchical learning, markedly improving model stability and accuracy, and providing scientific data support for e-commerce enterprise decision-making.

3.2. Neural network architecture

Optimizing neural network architecture is core to achieving high precision in e-commerce sales prediction models. RNN and its variants LSTM and GRU use gating mechanisms to capture temporal dependencies, effectively handling promotional fluctuations; CNN extracts local features from high-dimensional data to identify latent patterns; MLP integrates static non-sequential information such as product attributes [5]. Constructing hybrid models by fusing CNN, RNN, and MLP enables deep extraction of local patterns and long- and short-term features, strengthening nonlinear fitting capability. Strategies such as regularization, early stopping, and cross-validation significantly enhance generalization performance, providing scientific decision support for precise marketing and inventory management [6].

3.3. Feature engineering methods

Feature engineering is the key to extracting high-dimensional complex information and improving prediction performance in e-commerce sales models [7]. Traditional Pearson, Spearman, or Kendall correlation analyses have limitations in handling nonlinear relationships and noise. Researchers have introduced diverse deep learning approaches. For instance, Xue *et al.* used multi-objective random forests for feature selection [8].

Zha *et al.* combined CNN and LSTM for automatic feature extraction and temporal dynamic capture; Pan *et al.* employed self-attention mechanisms to strengthen feature dependency modeling; Cornelio *et al.* applied autoencoders to extract low-dimensional latent variables for nonlinear dimensionality reduction. This study integrates product basics, attributes, and promotional information, constructing features from five dimensions: summation, averaging, ratios, trends, and multi-attribute hybrids. By fusing traditional methods with deep learning's automatic modeling capability, the approach reduces manual intervention while significantly enhancing the model's fitting ability and prediction accuracy for complex patterns, providing an effective pathway for robust modeling [9].

4. Data collection and preprocessing

4.1. Data sources

Building an e-commerce sales prediction model relies on the integration of multi-source data. Sources include internal enterprise transaction databases, user logs, and third-party datasets, encompassing internal information such as product attributes and price fluctuations, as well as external contexts like industry trends and weather. To address noise and inconsistencies in raw data, preprocessing through data cleaning, outlier detection, and normalization is required, with time-series cross-validation used for dataset splitting. RNN, LSTM, and GRU models exhibit strong dependence on temporal structure, necessitating strict preservation of temporal logic during preprocessing. For massive data volumes, distributed frameworks such as Hadoop or Spark significantly improve processing efficiency [10]. Data quality and in-depth analysis of diverse features form the fundamental guarantee for model generalization capability and prediction accuracy.

4.2. Data cleaning

In deep learning-driven e-commerce sales prediction, data cleaning is the core step to ensure training quality and prediction accuracy. For missing values and anomalies in raw data, mean imputation, interpolation, and statistical methods such as box plots or Z-score are applied for correction to eliminate noise interference. To optimize convergence speed and training effectiveness of DNN and CNN, Z-score standardization or Min-Max normalization is used to eliminate scale differences among features. For time-series prediction tasks, continuity of timestamps must be ensured, and key temporal features such as months, holidays, and promotions are extracted to enhance the model's perception of sales fluctuations. Scientific partitioning into training, validation, and test sets provides structurally rigorous, high-quality data support for subsequent complex neural network architectures, ensuring excellent generalization capability and prediction stability [11].

4.3. Feature extraction

Feature extraction aims to mine core variables influencing sales. The preprocessing stage performs cleaning, imputation, and categorical encoding. Feature construction spans multiple dimensions: temporal (holidays and promotion cycles), product (price, category, and ratings), and user (behavior trajectories and membership levels). Additionally, user–product interaction features such as category preferences and conversion rates deepen the model's understanding of behavioral logic.

Under deep learning architectures, embedding layers vectorize discrete variables. CNN extracts local spatiotemporal features, while RNN or LSTM captures long-term temporal dependencies [5,11]. For unstructured data such as product descriptions, natural language processing techniques enable topic modeling and sentiment analysis, effectively expanding the feature input space.

To optimize feature quality, correlation analysis, chi-square tests, and random forest importance ranking are used to eliminate redundancy. Feature engineering must integrate algorithms with business logic to ensure features possess clear physical meaning. The optimized feature system lays a solid foundation for training, significantly improving sales prediction accuracy and stability [6].

5. Model construction

5.1. Model design

In e-commerce sales prediction model design, multi-dimensional time-series data serve as input, with future values as output. Architecturally, LSTM or Transformer captures long-range dependencies, while attention mechanisms enhance perception of key nodes; hybrid models incorporating CNN extract local features [7]. The loss function combines MSE with weighted loss to address uneven data distribution. Training involves missing value imputation and normalization preprocessing, with early stopping and learning rate adjustment to optimize convergence. Ultimately, MAE and RMSE metrics are used for evaluation, ensuring balance between accuracy and efficiency to effectively support inventory management [11].

5.2. Parameter settings

Model training parameters are systematically configured: initial learning rate is set to 0.0001 to balance convergence stability; input scale is 16 batches with 3 channels (224×224 images) over 150 epochs [12]. The optimization algorithm employs stochastic gradient descent (SGD) with momentum 0.7 and decay factor 0.00001 to enhance generalization [13]. Cross-entropy loss quantifies prediction deviation, providing clear guidance for model optimization [14]. This strategy achieves balance between computational efficiency and precision, ensuring robustness and high accuracy in the sales prediction model.

5.3. Training strategies

Training strategies must balance model complexity and generalization capability. Experiments use the Adam optimizer with parameters β1 = 0.9, β2 = 0.999, learning rate 1.5 × 104, batch size 8, and maximum 500 epochs. This configuration aims for smooth gradient descent and full capture of underlying data patterns.

The loss function adopts mean squared error (MSE), quantifying deviation via the mean of squared differences between multi-channel predictions and actual values [5]. This design is particularly suitable for continuous-output tasks such as sales prediction, enabling comprehensive and objective evaluation of fitting performance in multi-dimensional feature spaces. To address class imbalance, resampling, classifier adaptation, and multi-model ensemble strategies are integrated. GAN-based generative methods enhance data diversity, and sample weight adjustment effectively mitigates bias caused by data sparsity, improving prediction efficacy for low-frequency products.

In structural optimization, residual blocks suppress gradient vanishing and enhance deep learning capacity. Data augmentation strengthens information gain, while decoupled training separates classification and regression tasks, significantly improving tail-class recognition precision without loss of head information [14].

In summary, the training strategy achieves deep integration of parameter configuration and business logic through optimizer selection, MSE loss design, GAN-assisted augmentation, and structural improvements, providing robust system support for high-precision, high-efficiency e-commerce sales prediction.

6. Model optimization

6.1. Regularization techniques

Although neural networks offer strong fitting capability, they are prone to overfitting, impairing generalization performance. To improve model performance, this paper introduces early stopping as a regularization strategy. By dynamically monitoring validation error and automatically terminating training when the metric ceases to improve over consecutive iterations, the mechanism obtains the optimal parameter model [15]. This not only reduces computational overhead but also significantly suppresses overfitting to training data. In e-commerce sales prediction, reasonable configuration of monitoring windows and patience counts enables the model to maintain greater stability and prediction accuracy in complex market environments, providing robust support for decision-making.

6.2. Hyperparameter tuning

Hyperparameter configuration (e.g., number of neurons, learning rate, iteration steps) is crucial for handling high-dimensional, nonlinear e-commerce sales prediction. Since these parameters cannot be automatically adjusted, traditional manual tuning is highly subjective, while grid search faces extremely high computational costs and low efficiency in multi-dimensional complex models.

To balance accuracy and efficiency, this paper introduces Bayesian optimization and genetic optimization algorithms. The former constructs surrogate models and performs efficient sampling using historical information; the latter, despite global search capability, suffers from slow convergence and poor stability due to random walk characteristics. Comparative experiments on LightGBM demonstrate that Bayesian optimization achieves significantly superior prediction accuracy under the same number of searches [8].

This paper performs optimization for LightGBM, XGBoost, CatBoost, random forest, and DNN, using MSE as the objective. Results show that optimized LightGBM (MSE = 0.3448) performs best, with random forest and DNN errors 32.27% and 52.04% higher, respectively. Accordingly, the top three are selected as base learners for constructing the ensemble prediction model. For neural network tuning, experiments reveal that setting 60 hidden-layer neurons balances fitting capability and computational cost; a learning rate of 0.01 effectively balances training speed and accuracy; maximum iteration steps have relatively minor impact. Final parameters are configured as 60 neurons, 0.01 initial learning rate, and 7000 iterations to achieve optimal balance between performance and efficiency [15].

6.3. Ensemble learning

Ensemble learning improves prediction stability by fusing multiple base learners. Targeting the high-dimensional dynamics of e-commerce data, this paper uses evolutionary algorithms to adaptively optimize weights and constructs a fusion layer to provide feature inputs to the meta-learner. In feature selection, a binary adaptive differential evolution algorithm is designed to dynamically adjust search strategies. Given that random forest and XGBoost typically outperform logistic regression and support vector machines in comparisons, this paper constructs an adaptive evolutionary ensemble model [4]. By optimizing structure and feature strategies, the model effectively addresses hyperparameter tuning and generalization challenges, significantly enhancing sales prediction accuracy and robustness.

7. Experimental design and results analysis

7.1. Experimental design

The experiment systematically plans data governance, feature mining, and model tuning. Raw data first undergoes cleaning and standardization to ensure quality; key features are then extracted from multi-dimensional attributes, with text sentiment analysis introduced to quantify user comment polarity [6]. Architecturally, LSTM or Transformer captures deep temporal dependencies, with cross-validation, early stopping, and learning rate adjustment strategies optimizing the training process. Finally, MSE and MAE metrics evaluate prediction performance, with comparative experiments against LightGBM. This design aims to balance model accuracy and interpretability, providing robust scientific support for e-commerce business decisions.

7.2. Experimental environment

The experiment builds a stable and scalable platform based on distributed frameworks [9]. Computational resources are provided by multiple high-performance computing nodes. Data processing leverages Spark for offline computation and Spark Streaming for real-time stream processing to enable immediate model updates [9]. Storage architecture uses Hive for intermediate results, HBase for large-scale structured data supporting high-concurrency access, and Redis for caching configuration information and real-time analysis results [9]. Kafka facilitates asynchronous message passing between modules, ensuring system coordination. This environment effectively meets the demands of deep learning for massive data processing, laying a solid foundation for model iteration and deployment.

7.3. Results analysis

Performance evaluation shows that the model achieves a MAPE of 6.8% on the test set, demonstrating excellent prediction accuracy [5]. Comparative experiments indicate that this deep learning model outperforms traditional time-series models (ARIMA, SARIMA) and shallow machine learning models (random forest, XGBoost) in handling nonlinear trends and multivariate interactions [5,8]. The incorporation of attention mechanisms and feature importance analysis enables precise identification of key drivers such as promotions and comment sentiment. Through hyperparameter optimization and regularization strategies, generalization capability is effectively enhanced. Combined with real-time data stream processing, the model demonstrates significant practical value in inventory management and supply chain optimization.

8. Application case

8.1. Case selection

In constructing a deep learning-based e-commerce sales prediction model, case selection is critical to determining model performance and applicability. This study selects historical sales data from mainstream platforms such as Taobao, JD.com, and Pinduoduo from 2018 to 2023 as the experimental case, ensuring breadth and timeliness of samples. The dataset covers multidimensional features, including promotions, seasonal fluctuations, and new product launches, with preprocessing techniques (missing value imputation, outlier detection, and standardization) optimizing input quality. To enhance generalization of deep learning models, external environmental variables such as user behavior, weather, and economic indices are incorporated beyond product attributes. Comparative analysis across categories such as fast-moving consumer goods and 3C products identifies unique sales fluctuation patterns, providing a scientific basis for personalized model design [5]. Finally, partitioning into training, validation, and test sets ensures objectivity and reliability in model evaluation.

8.2. Model application

In e-commerce sales prediction, deep learning architectures achieve precision improvements through multi-granularity scanning and cascaded forests. In the multi-granularity scanning stage, the model receives feature sequences extracted by CNN and uses sliding windows to generate A-B+1 B-dimensional local feature subsets. Each subset is processed by two types of random forests to produce 2K(A-B+1)-dimensional class vectors, which are concatenated with original features to form enhanced inputs for the cascaded forest stage.

In the cascaded forest stage, layer-by-layer representation learning optimizes prediction performance. Each layer consists of random forests optimized by Gini coefficients and completely random forests with random split features to enhance model diversity and generalization. Training employs k-fold cross-validation (k = 5), with weighted prediction vectors generated according to specific weighting formulas. The cascaded structure dynamically extends based on prediction accuracy evaluation, ultimately selecting the best-performing layer as model output, providing high-precision predictions for e-commerce sales [5].

8.3. Performance evaluation

To more comprehensively assess model prediction capability, this paper introduces an improved weighted deep forest algorithm to optimize e-commerce sales prediction. Addressing differences in subtree prediction accuracy within deep forests, the algorithm constructs a weight allocation mechanism based on prediction accuracy. Subtree performance is evaluated via sales segmentation, and a weighting formula assigns greater influence to high-accuracy subtrees, effectively mitigating error amplification in traditional cascaded forests. Experiments demonstrate that this mechanism significantly enhances model stability and accuracy in handling high-dimensional nonlinear data, providing robust theoretical support and practical basis for precise sales prediction in complex e-commerce scenarios [5].

9. Conclusions and outlook

In constructing and optimizing a deep learning-based e-commerce sales prediction model, reasonable selection of model architecture and optimization strategies proves key to improving prediction accuracy. Comparative analysis of LSTM, GRU, Transformer, and hybrid models confirms that Transformer with attention mechanisms exhibits significant advantages in nonlinear fitting and long-term dependency capture. Preprocessing combined with integration of user behavior, environmental variables, and product attributes effectively strengthens model generalization. Training employs early stopping, cross-validation, and Adam optimizer with dynamic learning rates to ensure convergence while suppressing overfitting. Experimental results show that the optimized model outperforms traditional statistical models and shallow neural networks across MAE, RMSE, and MAPE metrics. Additionally, interpretability analysis based on feature importance provides scientific decision support for e-commerce inventory management and refined operations.

Performance evaluation should draw from medical prediction by introducing ROC curves and AUC metrics. For example, in KOA research, meniscus MRI models (AUC improved from 0.60 to 0.71) outperformed MOAKS and clinical information models. This provides a cross-disciplinary reference for fusing multidimensional data and enhancing stability in e-commerce prediction. Optimization should focus on feature quality. Given the insignificant improvement from joint models ($P > 0.05$), blindly stacking features easily introduces noise. Attention mechanisms should be used to automatically identify key sales drivers, balancing generalization performance and interpretability. Evaluation dimensions must be further expanded. In addition to DeLong tests for AUC differences, RMSE, MAE, and MAPE should be integrated for measurement. Combining visualization and interpretability algorithms can effectively reduce the "black-box" effect, supporting scientific business-layer decision-making. In summary, sales prediction requires coordination of data depth and evaluation rigor. Drawing on medical imaging research ideas and deeply exploring deep learning potential will continue to drive breakthroughs in prediction accuracy, providing robust support for supply chain optimization and operational decision-making.

In constructing a deep learning-based e-commerce sales prediction model, improvement measures focus on enhancing feature representation and architectural robustness. First, graph neural networks (GNN) are introduced to capture latent associations between users and products, synergizing with Transformer to strengthen global dependency modeling. Second, transfer learning addresses cold-start issues for new products, leveraging cross-category knowledge transfer to mitigate data sparsity. Simultaneously, deeper coupling of CNN and LSTM through adaptive feature fusion improves extraction precision of multidimensional spatiotemporal patterns. Furthermore, large language models (LLM) are explored for semantic mining of review data, continuously injecting feature gains into cascaded forest and other ensemble architectures to ensure prediction efficacy in dynamic market environments.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Xun K, Feng B, Chen Y, et al., 2022., Integrated Learning Method for Thymoma CT Image Prediction Combining Deep Learning and Improved Extreme Learning Machine. Computer Science, 2022(49): 455–460.

[2] Liu L, Chen X, Sun S, et al., 2022, Application of Deep Learning in Drug Activity Prediction Research. Progress in Biochemistry and Biophysics, 2022(49): 110–131.

[3] Li M, Sheng K, Xiong Z, et al., 2025, Research on the Application of Deep Learning in Plastic Processing. Plastics Industry, 53(462): 209–210.

[4] Zhang Y, Yu Q, Zhang L, 2024, User Purchase Intention Prediction based on Evolutionary Ensemble Learning. Application Research of Computers, 41(388): 54–60.

[5] Yuan R, Wei H, Fu Z, et al., 2025, Research on E-Commerce Product Sales Prediction Fusing CNN and WDF Models. Computer Engineering and Applications, 61(1065): 339–347.

[6] Huang H, Sheng W, 2022, Click Prediction Model based on Product Description Copy. Application Research of Computers, 39(370): 188–192.

[7] Zhou F, Chen X, Zhong T, 2022, A Survey of Deep Learning Techniques for Financial Technology. Computer Science, 2022(49): 20–36.

[8] Yan Y, Huang Q, Li N, 2023, Research on Feature Selection and Efficient Disease Early Warning based on

Optimized Ensemble Learning Models: A Case Study of Geriatric Depression and Anxiety. Data Analysis and Knowledge Discovery, 7(79): 78–92.

[9] Li N, Liu W, Meng F, 2022, Spatiotemporal Regional Economic Visualization Application Under Telecom Big Data Analysis. Journal of Beijing University of Aeronautics and Astronautics, 48(348): 84–92.

[10] Dang H, Wang Z, Zhao D, et al., 2024, Prediction of Plate Boundary Earthquakes in the Japan Subduction Zone based on Stacking Model Fusion Strategy. World Earthquake Engineering, 2024(40): 82–97.

[11] Guo Z, Ma B, Zhang S, et al., 2024, Research Progress and Technical Outlook of Deep Learning in Oil and Gas Production Prediction. Natural Gas Industry, 44(371): 94–104.

[12] Chen Z, Hong Y, Chen T, et al., 2025, Rapid Prediction Model for Tunnel Fire Smoke Layer Height and Back-Layering Length based on Transposed Convolutional Neural Networks. China Occupational Safety Science and Technology, 2025(21): 65–70.

[13] Gong C, Li Y, Liu J, et al., 2023, Rapid Automatic Adjustment Method for Large Power Grid Section Power based on Deep Learning. Automation of Electric Power Systems, 47(744): 186–195.

[14] Ren N, Fu Y, Wu Y, et al., 2022, Survey of Research on Imbalance Problems in Deep Learning for Object Detection. Computer Science and Exploration, 16(168): 5–25.

[15] Liu R, Luo K, Xing J, et al., 2022, A Neural Network-Based Combustion Model and its a Priori Validation. Combustion Science and Technology, 28(152): 75–81.

[16] Jiang K, Xie Y, Li M, et al., 2024, Research on Predicting Knee Osteoarthritis Onset using 3D Convolutional Neural Network Models based on Meniscus MRI. Chinese Journal of Magnetic Resonance Imaging, 15(140): 110–114+128.