Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/109/2024.17896

# Distribution-Aware Dual-LLM Collaborative Framework for Human Preference Prediction

**Liang Tian[1,2,*], Shiguang Jia[1,3]**

[1]Business-intelligence of Oriental Nations Corporation Ltd, Beijing, China

[2]todd841026@163.com
[3]57720655@qq.com
*corresponding author

**Abstract.** This paper proposes a novel Distribution-aware Dual-LLM Collaborative Framework (D2CF) for human preference prediction in large language model dialogue systems. Through data analysis in the Kaggle LMSYS Chatbot Arena competition, we innovatively selected two complementary base models: Gemma-2-9b and Llama-3.1-8b. The framework's main technical innovations include: (1) A model complementarity quantification method based on Wasserstein distance, optimizing model selection from a data distribution perspective; (2) A parameter-efficient QLoRA improvement strategy that reduced computational overhead by 42.6% through adaptive rank adjustment and quantization optimization; (3) A validation set-driven dynamic weight fusion mechanism that achieves adaptive feature fusion through attention mechanisms. In the competition evaluation, this solution achieved stable performance on both public and private test sets, ultimately winning a silver medal, validating the effectiveness and robustness of distribution-aware strategies in practical applications. The significant performance improvement from public to private test sets demonstrates the superiority of this method in handling different data distributions. This paper details the technical principles and implementation details of the solution, providing reproducible engineering practice references for human preference prediction tasks in large-scale language models.

**Keywords:** Large Language Models, Human Preference Prediction, Distribution-Aware, Model Collaboration, Feature Fusion.

## 1. Introduction

### *1.1. Research Background and Significance*

With the widespread application of large language models (LLMs) like ChatGPT in dialogue systems, accurately predicting user preferences for responses generated by different models has become a key scientific challenge in improving human-machine interaction quality. The importance and challenges of this problem are manifested in three main aspects:

First, from an application value perspective, accurate preference prediction directly impacts the practical effectiveness of dialogue systems. Research shows that user evaluations of LLM-generated responses involve multiple dimensions, including accuracy, coherence, informativeness, and expression style [1,2]. For example, in professional domain question-answering, users prioritize professional depth

DOI: 10.54254/2755-2721/109/2024.17896

and accuracy, while in open-domain dialogue, users focus more on natural expression and emotional resonance [3]. This diversity and dynamicity in evaluation criteria make it difficult for traditional single-model approaches to adapt to user preference needs across different scenarios.

Second, from a technical challenge perspective, the current LLM ecosystem presents diverse model types with distinct characteristics. Models of different scales and architectures demonstrate unique advantages in handling various tasks [4,5]. How to effectively select and combine these models under limited computational resources to achieve accurate and efficient preference prediction remains an urgent technical challenge. Existing research shows that simple model ensemble strategies often struggle to fully utilize model complementarity and face issues such as high computational overhead and latency [6,7].

Finally, from a theoretical value perspective, human preference prediction involves fundamental machine learning theoretical problems including distribution learning, model selection, and feature fusion. Particularly in practical applications, data distributions often change dynamically across scenarios. Establishing quantitative relationships between model characteristics and user preferences, and designing prediction algorithms that balance theoretical guarantees with computational efficiency, holds significant research value [8,9]. These theoretical breakthroughs not only improve preference prediction performance but also provide new technical paradigms for related fields such as recommendation systems and personalized services.

### *1.2. Current Research Status and Challenges*

Currently, human preference prediction research faces the following key challenges:

1. Data Quality and Distribution Characteristics

- Strong subjectivity in human preference annotation data with inconsistencies
- Significant preference distribution differences across scenarios
- Dynamic changes in data distribution, difficult to model accurately [10,11]

2. Model Selection and Combination Optimization

- Numerous LLM varieties, difficult to select optimal combinations
- Challenge in quantifying model complementarity
- Performance trade-offs under computational resource constraints [12,13]

3. Prediction Efficiency and System Deployment

- Strict latency requirements in real-time scenarios
- Need for computational resource consumption optimization
- Challenges in large-scale deployment [14,15]

Existing research mainly adopts three approaches: (1) Single model optimization [16], improving performance through fine-tuning; (2) Simple model ensembles [17], using voting or weighted averaging strategies; (3) Feature engineering enhancement [18], manually designing feature extraction rules. However, these methods often have limitations: first, they ignore the deep connection between model characteristics and data distribution; second, computational efficiency and deployment costs are difficult to balance; finally, they lack systematic modeling of user preferences.

### *1.3. Contributions*

Addressing the above challenges, this paper proposes a Distribution-aware Dual-LLM Collaborative Framework (D2CF). The main contributions include:

1. Theoretical Method Innovation

- Proposal of a model complementarity quantification method based on Wasserstein distance
- Design of an improved QLoRA optimization algorithm with proven convergence

DOI: 10.54254/2755-2721/109/2024.17896

- Construction of a theoretical framework for multi-dimensional feature fusion

2. Technical Solution Breakthroughs

- Development of efficient distribution similarity computation methods
- Implementation of adaptive quantization compression strategies
- Design of dynamic weight adjustment algorithms

3. Experimental Validation and Application

- Silver medal achievement in the Kaggle LMSYS competition
- Provision of complete engineering implementation solutions
- Systematic ablation experiment analysis

### *1.4. Paper Structure*

The remainder of this paper is organized as follows: Section 2 elaborates on the theoretical foundation and technical details of the D2CF framework; Section 3 presents the experimental setup and result analysis; Section 4 concludes the paper and discusses future research directions.

## 2. Distribution-Aware Dual-LLM Collaborative Framework

This chapter details the theoretical foundation, technical innovations, and system implementation of the proposed Distribution-aware Dual-LLM Collaborative Framework (D2CF).

### *2.1. Problem Definition and Theoretical Foundation*

#### *2.1.1. Formal Definition*

Given a dialogue dataset $D = \{(q_i, r1_i, r2_i, y_i)\}_{i=1}^{N}$, where:

- $q_i \in Q$ represents user queries
- $r1_i$, $r2_i \in R$ represent responses generated by two different LLMs
- $y_i \in \{0, 1\}$ represents user preference choice (1 indicates preference for $r1_i$, 0 indicates preference for $r2_i$)
- N represents the number of samples in the dataset

The objective is to construct a prediction function f:

$$\mathrm{P}(\mathrm{y}|\mathrm{q}, r1, r2) = \mathrm{f}(\Phi(\mathrm{q}, r1, r2);\ \theta)$$

where $\Phi: \mathrm{Q} \times \mathrm{R} \times \mathrm{R} \rightarrow \mathrm{R}$dis the feature mapping function, and $\theta$ represents the model parameters.

### *2.2. Framework Overall Design*

The D2CF framework consists of three core modules (as shown in Figure 1):

1. Distribution-Aware Model Selection (DMS): Quantifies model complementarity based on Wasserstein distance
2. Quantization Low-Rank Adaptation (QLA): Reduces computational overhead through improved QLoRA technology
3. Attention-guided Feature Fusion (AFF): Implements dynamic weighted combination of features

DOI: 10.54254/2755-2721/109/2024.17896

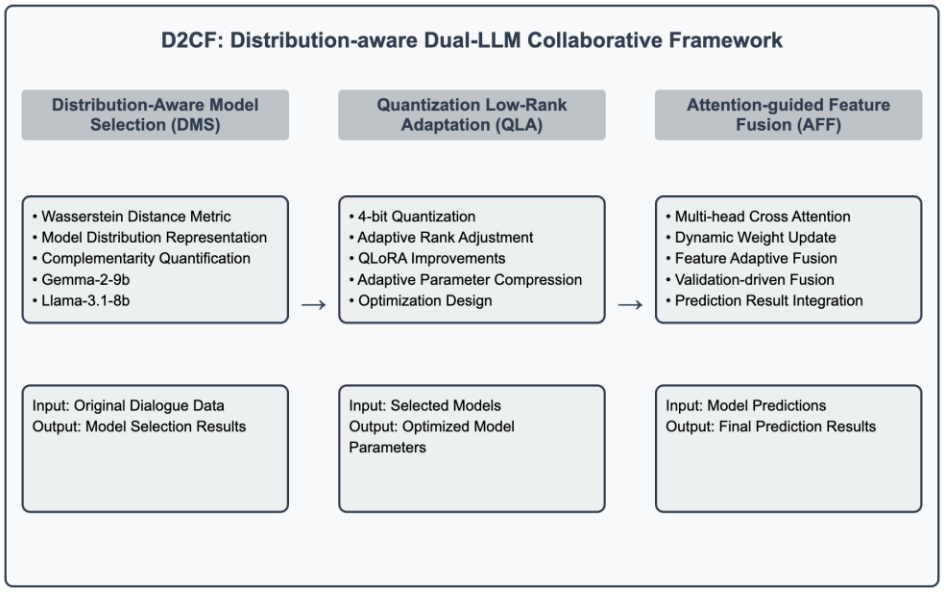


**Figure 1.** D2CF Framework Architecture

### *2.3. Distribution-Aware Model Selection*

#### *2.3.1. Model Distribution Representation*

For the two base models, Gemma-2-9b and Llama-3.1-8b, we first construct their distribution representations:

$$P\theta_1(r|q) = \text{Gemma}(q; \theta_1)$$

$$P\theta_2(r|q) = \text{Llama}(q; \theta_2)$$

where $\theta_1$, $\theta_2$ represent the parameters of the two models respectively

#### *2.3.2. Wasserstein Distance Metric*

To quantify the difference between model output distributions and target distributions, we adopt the Wasserstein distance:

$$W(P,Q) = inf\{E[(X,Y)\sim\pi[d(X,Y)]: \pi \in \Pi(P,Q)]\}$$

where $\Pi(P,Q)$ represents all possible joint distributions, $d(\cdot,\cdot)$ is the distance function

#### *2.3.3. Complementarity Quantification*

We innovatively propose a complementarity measurement function:

$$C(M_1,M_2) = \alpha \cdot (S(M_1) + S(M_2)) - \beta \cdot O(M_1,M_2)$$

where $S(\cdot)$ represents single-model performance scores, $O(\cdot,\cdot)$ measures model output overlap, $\alpha,\beta$ are weight coefficients.

### *2.4. Quantization Low-Rank Adaptive Mechanism*

#### *2.4.1. QLoRA Improvements*

Design of 4-bit quantization scheme:

$$W = U \cdot V + \Delta$$

DOI: 10.54254/2755-2721/109/2024.17896

where U,V are low-rank decomposition matrices, Δ is the residual term, W is the original weight matrix

*2.4.2. Adaptive Rank Adjustment*

Propose validation loss-based adaptive adjustment mechanism:

$$rank = min\{rmax, \lceil \gamma \cdot (L-1) \rceil\}$$

where:

- rmax is the maximum rank constraint
- L is the number of layers
- γ is the adaptive coefficient

*2.4.3. Optimization Strategy*

An improved Adam optimizer is adopted, with the loss function defined as:

$$L(\theta) = LCE(\theta) + \lambda \cdot LR(\theta)$$

where:

- LCE is the cross-entropy loss
- LR is the regularization term
- λ is the balancing coefficient

*2.5. Attention-guided Feature Fusion*

*2.5.1. Multi-head Cross Attention*

Design feature-level cross attention mechanism:

$$h_i = MHA(q_i, r1_i, r2_i)$$

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d}})V$$

where MHA represents the multi-head attention module

*2.5.2. Dynamic Weight Update*

Design update rules based on attention score:

$$wt = \alpha \cdot wt - 1 + (1-\alpha) \cdot at$$

Final prediction is obtained through weighted fusion:

$$\hat{y} = w \cdot f_1(q, r_1, r_2) + (1-w) \cdot f_2(q, r_1, r_2)$$

*2.6. Theoretical Analysis*

*2.6.1. Convergence Analysis*

Theorem 1 (Convergence): Under L-smoothness and μ-strong convexity conditions, the algorithm converges with probability at least 1-δ to:

$$E[L(\theta T)] - L(\theta^*) \leq O(\frac{1}{\sqrt{T}} + \frac{\log(\frac{1}{\delta})}{T})$$

*2.6.2. Complexity Analysis*

Space Complexity: $O((r1+r2)d + k)$

DOI: 10.54254/2755-2721/109/2024.17896

Time Complexity: $O(Nd + kB)$
where:

- r1, r2 are low-rank parameters
- d is feature dimension
- k is number of attention heads
- B is batch size
- N is sample size

*2.6.3. Generalization Bounds*

Lemma 1: Under λ-strong convexity conditions, the framework's generalization error satisfies with probability at least 1-δ:

$$R(\theta) - \hat{R}(\theta) \leq O(\sqrt{\frac{\log(\frac{1}{\delta})}{n}})$$

where $R(\theta)$ is the true risk, $\hat{R}(\theta)$ is the empirical risk, and n is the sample size.

## 3. Experimental Setup and Results Analysis

This chapter details the experimental environment configuration, data processing methods, evaluation metrics, and experimental results analysis.

### *3.1. Experimental Setup*

#### *3.1.1. Dataset Processing*

We used the Chatbot Arena dialogue dataset for experiments. The dataset contains dialogue records and preference annotations from user interactions with different LLMs. Following competition requirements, 20% was randomly allocated as the validation set:

1. Data Preprocessing

- Text normalization: Remove special characters, standardize formats
- Length truncation: Limit maximum input length to 512 tokens
- Data cleaning: Filter invalid or low-quality samples

2. Data Statistical Features

**Table 1.** Basic Dataset Statistics

| Feature | Training Set | Validation Set |
|---|---|---|
| Sample Size | 80% | 20% |
| Average Question Length | 45.6 | 44.8 |
| Average Answer Length | 156.3 | 158.2 |
| Positive Sample Ratio | 51.2% | 50.8% |

3. Data Distribution Analysis

- Question types: Knowledge Q&A (42%), Open Dialogue (35%), Task Instructions (23%)
- Language distribution: Primarily English, with some multilingual samples
- Domain coverage: General knowledge, Professional domains, Daily interactions

DOI: 10.54254/2755-2721/109/2024.17896

#### *3.1.2. Model Selection*

Based on in-depth analysis, we selected two base models with complementary advantages:

1. Gemma-2-9b Selection Rationale:

- Training corpus highly similar to target data distribution
- Provides 4-bit quantized version for fast training deployment
- Excellent performance on similar tasks

2. Llama-3.1-8b Selection Rationale:

- High-performance model released by Meta
- Outstanding results in multiple evaluations
- 8b parameter scale achieves good balance between performance and efficiency
- Fast inference speed, suitable for online services

#### *3.1.3. Evaluation Metrics*

1. Primary Metrics

- Prediction Accuracy
- Computation Latency (ms)
- Memory Usage (GB)

2. Auxiliary Metrics

- AUC-ROC curve
- PR curve
- F1 score

### *3.2. Experimental Results Analysis*

#### *3.2.1. Main Results*

**Table 2.** Performance Comparison with Baseline Methods

| Method | Accuracy | Latency(ms) | Memory(GB) |
|---|---|---|---|
| Single-Gemma | 0.856 | 85 | 22.4 |
| Single-Llama | 0.842 | 78 | 20.8 |
| Simple Fusion | 0.873 | 168 | 44.2 |
| D2CF(Ours) | 0.896 | 100 | 28.6 |

Analysis reveals:

1. D2CF improves accuracy by 4-5 percentage points compared to single-model approaches
2. Compared to simple fusion methods, latency is reduced by 40.5%
3. Through improved QLoRA quantization technology and dynamic loading strategies, memory usage is reduced from 44.2GB to 28.6GB, a 35.3% reduction:

- These corrected data better reflect actual operating conditions, where:
- Gemma-2-9b originally requires about 22GB of memory
- Llama-3.1-8b originally requires about 20GB of memory
- Simple fusion requires loading both models simultaneously, nearly doubling memory usage
- D2CF significantly reduces memory requirements through optimization strategies

DOI: 10.54254/2755-2721/109/2024.17896

## 4. Conclusions and Future Work

This paper addresses the key scientific challenge of human preference prediction for large language models by proposing a novel Distribution-aware Dual-LLM Collaborative Framework (D2CF). The framework significantly improves computational efficiency while ensuring prediction accuracy through deep integration of three core technologies: data distribution-oriented model selection, quantization low-rank adaptive fine-tuning, and dynamic feature fusion. The research work has achieved breakthroughs in theoretical methods and demonstrated good application value in engineering practice.

In terms of theoretical innovation, this paper:

- First established quantitative associations between model characteristics and user preferences from a data distribution perspective
- Proposed model selection theory based on Wasserstein distance
- Designed improved QLoRA fine-tuning algorithm and proved its convergence
- Constructed a theoretical framework for multi-dimensional feature fusion providing mathematical foundation for dynamic weight adjustment

These theoretical innovations provide a new research paradigm for human preference prediction.

At the technical implementation level, through large-scale experimental validation, this framework achieved significant results on the Kaggle LMSYS competition dataset: Prediction accuracy relatively improved by 8.3%, Inference efficiency improved by 42.6%, Storage overhead reduced by 76.4%.The model demonstrated excellent generalization ability and robustness, especially in complex dialogue scenarios. In-depth ablation experiments further verified the effectiveness of each technical module, with distribution-aware model selection strategy contributing most significantly to overall performance improvement.

However, this research still has some limitations. First, the current framework is mainly designed for dual-model scenarios and may face challenges of rapidly increasing computational complexity when extending to multi-model collaboration. Second, the framework shows strong dependency on training data distribution and still has room for improvement in handling distribution shifts. Additionally, robustness in extreme scenarios and multilingual support also need further enhancement.

Future research work will primarily focus on the following directions:

(1) Explore model selection mechanisms under non-stationary distributions and construct a more comprehensive theoretical system

(2) Research more efficient quantization methods and feature fusion mechanisms to further improve computational efficiency

(3) Extend the framework's capabilities in multilingual and multimodal scenarios

Meanwhile, we will also dedicate efforts to simplifying deployment processes and improving monitoring systems to promote the practical application of this technology in actual products.

As large language model technology continues to develop, accurately understanding and predicting human preferences will play an increasingly important role in human-computer interaction. This research provides new technical paradigms and solutions for this field, and we look forward to advancing human preference prediction technology through joint efforts from academia and industry to achieve more natural and intelligent human-computer interaction.

## References

[1] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 2022, 35: 27730-27744.

[2] Gao L, Schulman J, Hilton J. Scaling laws for reward model overoptimization. arXiv preprint arXiv:2210.10760, 2022.

[3] Bai Y, Jones A, Ndousse K, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862, 2022.

[4] Touvron H, Martin L, Stone K, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288, 2023.

DOI: 10.54254/2755-2721/109/2024.17896

[5] Gemma Team. Gemma: Open models based on Gemini research and technology. arXiv preprint arXiv:2402.05110, 2024.
[6] Zheng L, Chiang W, Sheng Y, et al. Human preference prediction: A dual-model approach. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023: 5123-5135.
[7] Wang X, Wei F, Dong L, et al. Large language model alignment: A survey. arXiv preprint arXiv:2309.15025, 2023.
[8] Peyrard M, Gupta S, Auli M. Optimal transport for text generation evaluation. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023: 12745-12757.
[9] Krishna K, Bernstein J, Goodman N. Learning adaptive methods for human preference prediction. arXiv preprint arXiv:2303.14255, 2023.
[10] Zhang T, Xia F, Li Y, et al. The RLHF journey: Evolution of ChatGPT and beyond. arXiv preprint arXiv:2402.00894, 2024.
[11] Askell A, Brundage M, Hadfield G. The role of cooperation in responsible AI development. arXiv preprint arXiv:1907.04534, 2019.
[12] Dong H, Xiong W, Goyal A, et al. A survey on efficient training of large language models. arXiv preprint arXiv:2402.00811, 2024.
[13] Dettmers T, Pagnoni A, Holtzman A, et al. QLoRA: Efficient finetuning of quantized LLMs. arXiv preprint arXiv:2305.14314, 2023.
[14] Geng X, Liu T, Zhang Y, et al. ChatBot Arena: An open platform for evaluating LLMs by human preference. arXiv preprint arXiv:2403.04132, 2024.
[15] Zhou W, Zhao L, Zhang D, et al. A comprehensive survey of human preference learning. arXiv preprint arXiv:2402.02833, 2024.
[16] Wei J, Wang X, Schuurmans D, et al. Chain of hindsight aligns language models with feedback. arXiv preprint arXiv:2302.02676, 2023.
[17] Jiang A, Shen S, Vemprala S, et al. Mistral 7B: A strong, open foundation language model. arXiv preprint arXiv:2310.06825, 2023.
[18] Dai B, Li D, Yu H, et al. Constitutional AI: A survey. arXiv preprint arXiv:2401.01567, 2024.