Explanation of Revisions: CodePMP: Scalable Preference Model Pretraining for Large Language Model Reasoning

Anonymous ACL submission

1 Overview

This document explains the revisions made to our manuscript "CodePMP: Scalable Preference Model Pretraining for Large Language Model Reasoning" since the previous submission. The previous submission was desk rejected due to incomplete submission form fields rather than substantive reviewer feedback. Therefore, the current resubmission incorporates targeted improvements to strengthen the manuscript without major structural changes.

2 Summary of Changes

The revisions focus on three main areas: (1) enhanced experimental analysis and details, (2) updated related work with recent concurrent research, and (3) additional computational cost information for reproducibility.

3 Detailed Changes

3.1 Enhanced Experimental Analysis

We have expanded our experimental analysis to provide more comprehensive insights into the CodePMP approach:

- **Cross-architecture evaluation**: Added evaluation results on Gemma2 and Llama3.2 model families (Section 4.5.3) to demonstrate the broad applicability of CodePMP across different model architectures.
- **Larger backbone model analysis**: Included experiments with Qwen2-72B (Section 4.5.4) to validate CodePMP's effectiveness at larger model scales.
- **Advanced generator evaluation**: Extended evaluation to more powerful generators including Qwen2-Math-7B-Instruct and Qwen2.5-32B-Instruct (Section 4.5.5) to demonstrate CodePMP's robustness across different generation capabilities.
- **General RM benchmark evaluation**: Added comprehensive evaluation on RMBench (Section 4.5.6) to assess CodePMP's applicability beyond reasoning tasks, covering summarization, chat quality, and safety.
- **Expanded ablation studies**: Enhanced ablation studies with additional analysis of loss function components (Section 4.4.2) and comprehensive data diversity analysis (Appendix Section A.6).

3.2 Updated Related Work

We have updated the related work section to include recent concurrent research in preference modeling:

- **Concurrent work integration**: Added citation and discussion of WorldPM (1) as concurrent work that also explores scaling human preference modeling, properly distinguishing it from prior work to maintain academic integrity.
- **Contextualized positioning**: Better positioned our work within the broader landscape of preference model pretraining approaches, highlighting the unique focus on code-derived preference data for reasoning tasks.

3.3 Computational Cost Information

We have added detailed computational cost information to improve reproducibility:

- **Training resource requirements**: Added comprehensive computational cost information (Appendix A.1) specifying that Qwen2-7B CodePMP training requires 128 H800 GPUs for 20 hours, while Qwen2-1.5B training requires 64 H800 GPUs for 12 hours.
- **Structured presentation**: Organized computational cost information in both textual description and tabular format for easy reference.

- **Updated appendix title**: Modified Appendix A title from "Hyperparameters" to "Hyperparameters and Computational Cost" to reflect the expanded content.

4 Technical Improvements

Beyond the main content changes, we have also made several technical improvements:

- **Enhanced data analysis**: Provided more detailed analysis of synthetic data quality and diversity, including n-gram distribution analysis and embedding space visualization.
- **Extended experimental scope**: Expanded evaluation to include more diverse model families and generation scenarios to strengthen the generalization claims.
- **Improved clarity**: Enhanced presentation of experimental results with better figure organization and clearer explanations of key findings.

5 Conclusion

The revisions maintain the core contributions and methodology of the original submission while significantly strengthening the experimental validation and positioning within current research. The changes address potential concerns about generalizability, computational requirements, and comprehensive evaluation across different scenarios. These improvements enhance the manuscript's contribution to the field without altering its fundamental approach or conclusions.

The resubmission is complete with all submission form fields properly filled, addressing the technical issue that led to the previous desk rejection.

References

[1] Binghai Wang, Runji Lin, Keming Lu, Le Yu, Zhenru Zhang, Fei Huang, Chujie Zheng, Kai Dang, Yang Fan, Xingzhang Ren, An Yang, Binyuan Hui, Dayiheng Liu, Tao Gui, Qi Zhang, Xuanjing Huang, Yu-Gang Jiang, Bowen Yu, Jingren Zhou, and Junyang Lin. Worldpm: Scaling human preference modeling. *arXiv preprint arXiv:2505.10527*, 2025.