NOSE: Neural Olfactory Sensing and Evaluation

Wenye Xiong
ShanghaiTech University
2023533141
xiongwenye@shanghaitech.edu.cn

Jun Wang
ShanghaiTech University
2023511040
wangjun2023@shanghaitech.edu.cn

Ge Gao
ShanghaiTech University
2023533052
gaoge2023@shanghaitech.edu.cn

Abstract

With the rapid advancement of artificial intelligence, AI systems have demonstrated high efficiency in processing visual, auditory, and other perceptual information. However, olfactory information modeling remains underdeveloped due to the lack of large-scale datasets annotated with olfactory labels from human experts. MoLFormer, a Transformer-based model of chemical structure, has proved to have the ability to encode representations highly aligned with human olfactory perception on general chemical structure. In this work, we fine-tuned MoLFormer on the GS-LF dataset, resulting in improved stability and accuracy in its ability to distinguish and rate different odors. Experimental results show that, in certain downstream tasks, MoLFormer even outperforms OpenPOM, the current state-of-the-art model in olfactory prediction, highlighting its strong potential for machine olfactory perception.

1 Introduction

The intersection of artificial intelligence and human sensory perception offers profound scientific and industrial opportunities. While fields like machine vision and audition have been revolutionized by deep learning, machine olfaction—-the ability of AI to predict and interpret smells from molecular structures-—remains a formidable challenge. This difficulty is largely rooted in the complex, high-dimensional nature of chemical-perceptual space and a historical scarcity of large-scale, systematically labeled olfactory datasets. A promising direction to overcome this data limitation is to leverage the powerful representations learned by large models pre-trained on vast unlabeled datasets of molecular structures.[5]

Recent work has validated this approach, demonstrating that MoLFormer, a Transformer-based model pre-trained on chemical structures, can generate representations that are remarkably aligned with human olfactory perception even in a zero-shot setting [4]. However, despite this strong baseline alignment, its performance across standard evaluation metrics still falls short of specialized, state-of-the-art models like OpenPOM [3, 5], which was specifically designed for olfactory prediction. This raises a critical question: can the generalized knowledge of a large chemical model be adapted to achieve state-of-the-art performance in the specialized domain of olfaction?

In this work, we argue that this performance gap can be closed through targeted fine-tuning. We hypothesize that the rich, general-purpose representations of the pre-trained MoLFormer serve as an excellent foundation that, with modest task-specific training, can be refined to excel at olfactory

Introduction To Machine Learning (25 Spring).

prediction. To investigate this, we fine-tune the public MoLFormer model on the GS-LF dataset, a benchmark collection of molecules annotated with human olfactory ratings.

Our experiments confirm this hypothesis. The fine-tuned MoLFormer achieves results that are highly competitive with OpenPOM. Notably, on several key metrics for odor classification and pleasantness prediction, our model matches or even surpasses the performance of the current state-of-the-art. These findings demonstrate that fine-tuning pre-trained chemical models is a highly effective strategy for developing top-tier models for machine olfaction.

2 Methodology

2.1 Datasets

Our datasets are all publicly available from the Pyrfume Project.[1]

Leffingwell-Goodscent (GS-LF): We employ a curated and merged version of the Goodscents and Leffingwell datasets, provided by OpenPOM.[3] This dataset contains 4983 molecules with 138 expert-labeled descriptors (e.g. creamy, grassy), where each odorant may be linked to multiple descriptors.

Keller-2016[2]: This dataset contains ratings of 480 structurally and perceptually diverse molecules by 55 human participants, evaluating 23 descriptors. In this work, we perform an additional binarization step to complete the classification task.

2.2 Related Work

2.2.1 MoLFormer

MoLFormer is built upon the Transformer architecture and incorporates key innovations such as a linear attention mechanism and Rotary Positional Embeddings[4]. By leveraging the computational efficiency of linear attention and the compactness of SMILES representations, MoLFormer achieves highly efficient training performance in large-scale chemical language modeling tasks[4]. Benefiting from the incorporation of Rotary Positional Embedding, the model effectively encodes relative positional information, which is inherently aligned with the nature of molecular representation learning.

2.2.2 OpenPOM

OpenPOM[3] is the publicly available version of the Principal Odor Map (POM) model. The POM is a high-dimensional continuous space that preserves hierarchical and distance relationships in odor perception. OpenPOM has been trained on the GS-LF datasets[3] to predict odor perceptual labels. This training process allows it to capture the intricate patterns between molecular structures and the resulting odor perceptions. It is the penultimate layer of the graph neural networks (GNNs), specifically message-passing neural networks (MPNNs), to represent molecules as graphs, learning features of atoms and bonds.

In this work, we extract representations from the penultimate layer of OpenPOM. It is then used for downstream tasks, such as classification and rating, on the Keller-2016[2] dataset, comparing it with MoLFormer and fine-tuned MoLFormer as a state-of-the-art baseline model.

2.3 Fine-tuning MoLFormer

The pre-trained MoLFormer architecture is equipped with multiple output heads, each designed to process the model's molecular embeddings for a different downstream task. For our work, we leverage these built-in heads to adapt the model for olfactory prediction.

Our fine-tuning procedure follows a certain methodology: As shown in Figure 1, the parameters of the selected output head are kept frozen. The primary training objective is to adjust the weights of the MoLFormer's main encoder body with the GS-LF dataset, where each molecule's SMILES string is first converted into a sequence of token IDs by the `MolTranBertTokenizer`. With the output head frozen, only the parameters of the Transformer body are updated by minimizing a task-specific mol

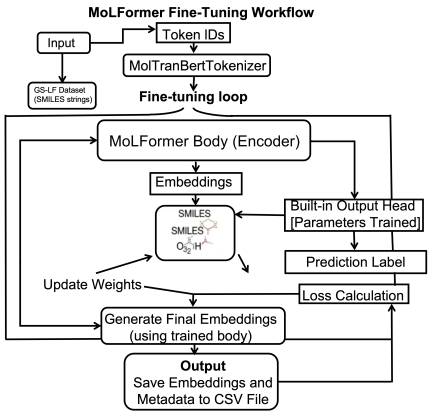


Figure 1: Pipeline of the fine-tuning workflow.

loss function. This approach allows the model's core representational capabilities to be refined for olfactory perception.

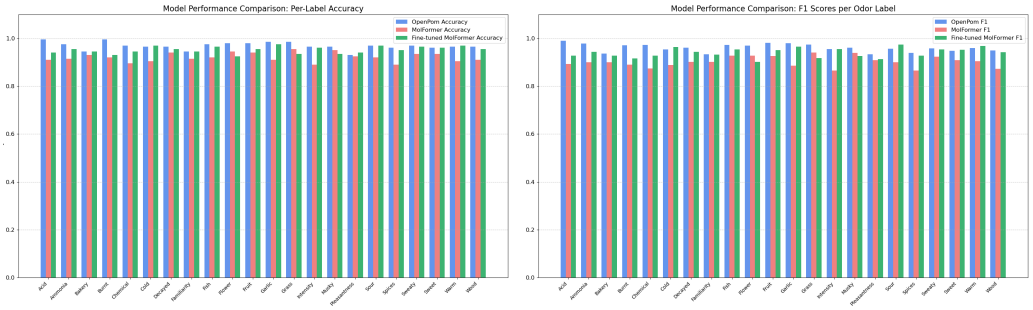
To generate embeddings for downstream analysis, we passed each molecule through the fine-tuned model body. While embeddings can be extracted from various hidden layers to capture different levels of molecular representation, we used the output from the final hidden layer for our primary tasks. These resulting high-dimensional embeddings, along with their corresponding metadata (e.g., CID, subject ID, and target perceptual values), were systematically saved to CSV files for subsequent machine learning experiments.

3 Result

In this section, we evaluate the performance of pre-trained MoLFormer, fine-tuned MoLFormer, and OpenPOM in several downstream tasks. Specifically, we focus on their ability to predict expert-assigned labels from odorants and to predict pleasantness scores provided by human participants.

3.1 Expert-assigned labels classification

To evaluate the performance of MoLFormer(both pre-trained and fine-tuned) in predicting expert-assigned labels for odorants, we implemented a linear mapping from the representations to the odorant representations extracted from the Keller-2016 dataset. First, the dimensionality of the extracted representations is reduced to 20 using PCA, followed by z-scoring of each feature. The parameters are chosen based on the number of descriptors in the Keller dataset. Then, we train individual logistic regression models for each descriptor. Both F1 Score and Accuracy are evaluated for each descriptor.



(a) Per-Label Accuracy (b) F1 Scores per Odor Label

Figure 2: Accuracy and F1 Score Comparison

As shown in Figure 2, OpenPom acts as a strong performance baseline, while the pre-trained MoLFormer underperforms. The central finding is that our fine-tuning process yields a substantial improvement, significantly elevating both accuracy and F1 scores across all odor labels to become highly competitive with the state-of-the-art. Notably, for specific labels such as "Decayed," "Garlic," and "Spices," the fine-tuned MoLFormer matches or even slightly surpasses the performance of OpenPom.

3.2 Pleasantness Rating Scores Prediction

To evaluate the ability of different models to predict pleasantness rating scores with respect to each descriptor, we train separate linear regression models with regularization applied using the Lasso penalty. Once again, the dimensionality of the extracted representations is reduced to 20 using PCA followed by z-scoring of each feature.

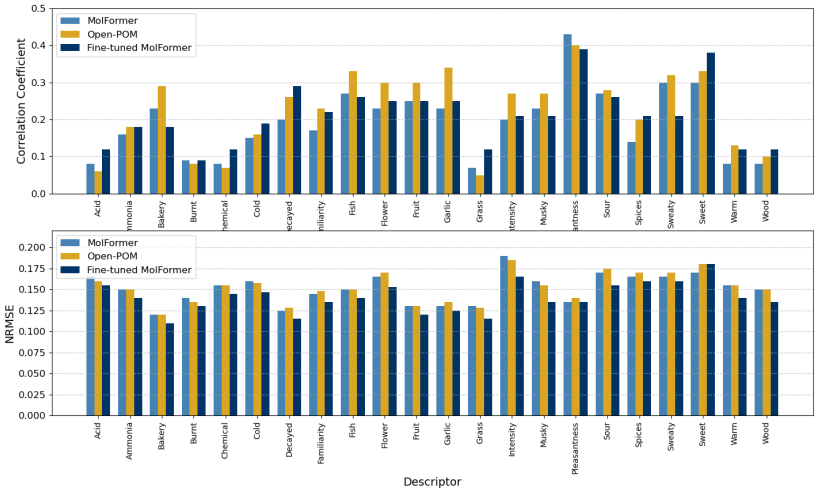


Figure 3: Performance of the models to predict continuous ratings averaged across all perceptual descriptors. We compute the average Pearson correlation coefficient (CC) and normalized root mean squared error (NRMSE) across all descriptors.

The results of these experiments are shown in Figure 3. Overall, none of the models exhibit a high correlation. Nevertheless, MoLFormer slightly underperforms compared to OpenPOM, and Fine-tuned MoLFormer achieves a better performance than the pre-trained one, close to or even outperforming the OpenPOM baseline. This demonstrates that while the general-purpose, pre-trained molecular representations are not perfectly suited for this specific olfactory task, the fine-tuning process effectively adapts the model, allowing it to achieve a competitive, state-of-the-art performance.

4 Discussion

Our contribution lies in demonstrating the significant performance gains achieved by fine-tuning the MoLFormer model for olfactory prediction. By adapting the pre-trained model to task-specific data, we elevated its performance in both expert-assigned odor label classification and human-rated pleasantness prediction, enabling it to match and in some cases even surpass the SOTA OpenPOM baseline.

Limitations: Our current methodology relies exclusively on chemical structure. This approach neglects other critical dimensions that influence the perception of smell, such as the physical properties of odorants (e.g. volatility) and their biological interactions (e.g. the binding affinity with human olfactory receptors). Incorporating these multifaceted features represents a crucial next step toward developing more holistic and accurate models of human olfaction.

References

[1] Jason B. Castro, Travis J. Gould, Robert Pellegrino, Zhiwei Liang, Liyah A. Coleman, Famesh Patel, Derek S. Wallace, Tanushri Bhatnagar, Joel D. Mainland, and Richard C. Gerkin. Pyrfume: A window to the world's olfactory data. *bioRxiv*, 2022. doi: 10.1101/2022.09.08.507170. URL `https://www.biorxiv.org/content/early/2022/09/12/2022.09.08.507170`.

[2] Andreas Keller and Leslie B. Vosshall. Olfactory perception of chemically diverse molecules. *bioRxiv*, 2016. doi: 10.1101/049999. URL `https://www.biorxiv.org/content/early/2016/04/23/049999`.

[3] Brian K. Lee, Emily J. Mayhew, Benjamin Sanchez-Lengeling, Jennifer N. Wei, Wesley W. Qian, Kelsie A. Little, Matthew Andres, Britney B. Nguyen, Theresa Moloy, Jacob Yasonik, Jane K. Parker, Richard C. Gerkin, Joel D. Mainland, and Alexander B. Wiltschko. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023. doi: 10.1126/science.ade4401. URL `https://www.science.org/doi/abs/10.1126/science.ade4401`.

[4] Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4(12):1256–1264, 2022. doi: 10.1038/s42256-022-00580-7.

[5] Farzaneh Taleb, Miguel Vasco, Antônio H. Ribeiro, Må rten Björkman, and Danica Kragic. Can transformers smell like humans? In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 72032–72060. Curran Associates, Inc., 2024. URL `https://proceedings.neurips.cc/paper_files/paper/2024/file/848350afc07c02edea5be06d3390d865-Paper-Conference.pdf`.