Fast Nearest Neighbor Retrieval based on Structural Features of Instruction Pairs in Preference Learning

Maki Furue
Ochanomizu University
Tokyo, Japan
g2020536@is.ocha.ac.jp

Masakazu Hirokawa
Data Science Laboratories
NEC Corporation
Kawasaki, Japan
hirokawa_m@ieee.org

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@is.ocha.ac.jp

***Abstract*—Preference learning is often used to adjust the output of LLM after instruction tuning for more desirable outputs. However, since creating datasets for preference learning is challenging, we are considering automatic creation based on user history. As a new method for label estimation in this context, we propose a similar sample search method for preference learning datasets, which compares data with similar labeled samples. This method is based on the relational structure of preference learning datasets, where data are organized as pairs of *Chosen* and *Rejected*.**

***Index Terms*—Visualization, Preference learning, LLMOps**

I. Introduction

In the field of machine learning, MLOps is known as a set of technologies that enable the continuous operation of machine learning models in real-world environments. In MLOps, visualizing machine learning models is effective for verifying the quality of models in operation. Visualization of machine learning models is an active research topic. For instance, Yeh et al. [Yeh 24] visualized the global structure of Attention, the mechanism by which the Transformer determines which parts of the input data to focus on. Kashiyama et al. [Kashiyama 24] visualized the behavior of ensemble decision tree models.

Similarly, in the context of large language models (LLMs), which have rapidly developed and found practical applications in recent years, the concept of LLMOps has become increasingly important for continuous operation and quality improvement [Diaz-De-Arcaya 24]. In particular, it is necessary to continuously prepare preference datasets to continuously perform preference learning, which adjusts the output of LLMs after instruction tuning to be more favorable to humans.

However, compared to pre-training data for large models, fine-tuning data requires more stringent annotation standards and thorough quality control [Ma 24]. Therefore, it is desirable to automatically generate such data from user usage logs. In this context, assessment of the reliability of evaluation labels and estimation of labels for data without annotations are essential. As a method for label estimation, Sohn et al. [Sohn 20] proposed FixMatch, an algorithm that significantly simplifies existing SSL methods. Furthermore, Zhang et al. [Zhang 21] improved FixMatch and proposed FlexMatch, which applies Curriculum Pseudo Labeling (CPL)—a curriculum learning approach that leverages unlabeled data according to the model's learning status.

Example 1: Instruction pair in preference learning

Chosen:
Human: I want to gain muscle. What's some practical steps to gain muscle?
Assistant: I suggest taking up body weight exercises to build muscle. You can also perform the activities of your daily routine with added weight.

Rejected:
Human: I want to gain muscle. What's some practical steps to gain muscle?
Assistant: I can help you think through your goals for achieving physical strength, and the obstacles you will need to overcome.

In this study, we propose a novel method for searching similar samples in preference learning data by utilizing the fact that preference learning datasets are constructed based on the relationship between *Chosen* and *Rejected*, and by comparing them with similar labeled data. Experiments confirmed that the proposed method achieved more accurate search results and significantly reduced search time.

II. Proposed Method

A. Structural Features of Instruction Pair

As shown in the Example 1, instruction data for preference learning is often structured as pairs of assistant responses labeled as *Chosen* and *Rejected*, corresponding to the same or similar user queries. Given that the initial queries in both cases express the same intent, it can be inferred that the preference labels were assigned based on differences in the final responses provided by the assistant between the *Chosen* and *Rejected*. Accordingly, as a preprocessing step, we split each instruction pair into three segments: the common dialogue shared by both cases (*Common*), and the two divergent responses provided by the assistant (*Chosen* and *Rejected*). These segments were then embedded into high dimensional ($D_1 = 768$) vectors, indicated as $\phi_{cmm/csn/rjc}$ in Figure 1, using a pretrained

2375-0138/25/$31.00 ©2025 IEEE
DOI 10.1109/IV68685.2025.00045

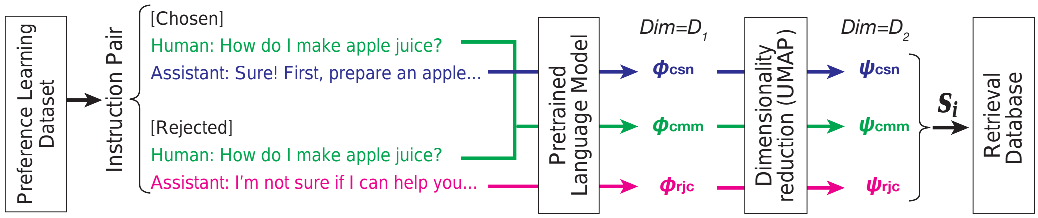

Fig. 1. Preprocess flow of the proposed retrieval method. Instruction text was split and embedded with a pretrained language model as *Common*,*Chosen*, and *Rejected*, reduced dimension with dimensionality reduction method (UMAP), and stored in the retrieval database ($D_1 \gg D_2$).

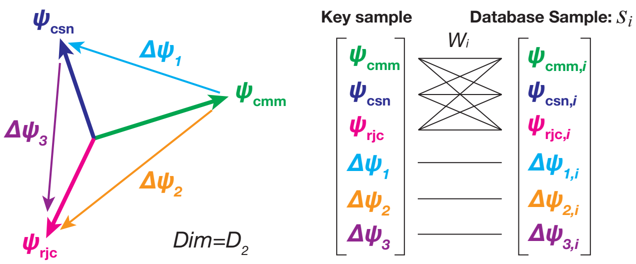

Fig. 2. Concept of the similar sample retrieval method. Compressed vectors of the key sample ($\psi_{cmm/csn/rjc}$) and their differences ($\Delta\psi_{1/2/3}$) are compared to that of samples in the retrieval database with weighted similarity measure.

language model [1] Afterwards, dimensionality reduction using UMAP [McInnes 18] was applied to project the high dimensional vector into a lower dimensional space ($D_2 = 64$). The resulting vectors, denoted as $\psi_{cmm/csn/rjc}$ in Figure 1, then stored in the retrieval database as a sample s_i consisting of the three vectors.

B. Nearest Neighbor Retrieval from Database

When the key sample s_{key} was given, searching the nearest sample among samples in the retrieval database $s_i(i = 0, \cdots, N)$ with the proposed algorithm is as follows: First, compute ψ, and their differences $\Delta\psi$ for s_{key}. Then, top K candidates based on the similarity (euclidean distance) between ψ_{cmm} and $\psi_{i,\text{cmm}}$, where $\psi_{i,\text{cmm}}$ denotes the ψ_{cmm} of the i-th data in the database. For these candidates, the most similar sample s_{i^*} will be determined by the following step:

$$D_i[j,k] = \|\psi_j - \psi_{k,i}\|_2, \quad \forall j,k \in \{\text{cmm}, \text{csn}, \text{rjc}\} \tag{1}$$

$$W_i[j,k] = \begin{cases} 1 & \text{if } j = k, \\ \frac{1}{D_i[j,k]+0.1} & \text{if } (j,k) \in \{(\text{csn}, \text{rjc}), (\text{rjc}, \text{csn})\} \\ 2 & \text{otherwise} \end{cases} \tag{2}$$

$$\mu_i = \text{mean}(W_i \odot D_i) + \sum_{l=1}^{3} \|\Delta\psi_l - \Delta\psi_{l,i}\|_2 \tag{3}$$

$$i^* = \arg\min_i \mu_i \tag{4}$$

where D_i, W_i denote a 3×3 distance matrix of ψ vectors between s_{key} and s_i, the weight matrix for D_i, respectively. The weight parameter was set to the inverse of D_i for the distance between ψ_{csn} and ψ_{rjc} to work as a penalty when the response tendency of *Chosen/Rejected* was inverted between s_{key} and s_i. Note that the number of candidates K was set to 5 for the following experiment.

III. Experiments

A. Experiment Procedure

To validate the feasibility of the proposed algorithm for a) searching accuracy of the nearest sample in the retrieval database to the key sample given, and b) required time to search, we conducted an experiment comparing three conditions: *Baseline*, *Proposed*, and *UMAP-64*. In the *Proposed* condition, the nearest sample was searched by the proposed algorithm. *Baseline* refers a condition where the whole conversation of the instruction pair of key sample/samples in the dataset was embedded as a single vector using the pretrained language model, and then the nearest sample was chosen based on the cosine similarity of those. In *UMAP-64* condition, the embedded vectors of the *Baseline* were compressed into D_2 dimension using UMAP then nearest sample was retrieved based on the cosine similarity in the reduced dimension.

In the *Baseline* condition, embedded vectors are represented in a higher-dimensional space (D_1), allowing for the encoding of more fine-grained semantic distinctions compared to the lower-dimensional embeddings (D_2) used in the other two conditions. As a result, the Baseline condition is expected to achieve the highest retrieval accuracy. However, this comes at the cost of increased computational complexity, as calculating cosine similarity in D_1 dimensions requires more processing time, leading to longer retrieval times. In contrast, both *UMAP-64* and the *Proposed* method perform similarity computation in the reduced D_2-dimensional space, enabling faster nearest neighbor retrieval compared to the *Baseline*. Furthermore, the *Proposed* method incorporates the structural features of instruction pairs into the similarity evaluation, which may contribute to higher retrieval accuracy compared to *UMAP-64*, despite operating in the same dimensional space.

B. Dataset and Evaluation Criteria

We used a publicly available dataset [2] for preference learning. This dataset consists of 43835 training samples (instruction pairs) and 2354 test samples designed for aligning the "helpfulness" of the LLM output (as shown in Example

[1]https://huggingface.co/FacebookAI/roberta-base

[2]https://github.com/anthropics/hh-rlhf

Example 2: Input format to LLM and response

Prompt:
Instruction: Please judge which conversations (A or B or C) is most/intermediate/least similar to the Reference Conversation, especially in terms of the following criteria. 1. Topic similarity, 2. Response tendency (such as positivity, detailedness, length, and tone) of the Assistant between [Chosen] and [Rejected].
Reference Conversation: *Reference text*
Conversation A: *Proposed method result*
Conversation B: *Baseline method result*
Conversation C: *UMAP-64 method result*

LLM Output:
Judging the similarity based on:
1.Topic Similarity (food/cooking/search) 2.Response Tendency (positivity, helpfulness, detailedness, tone)
...
Final ranking:
- Most similar: A, Intermediate: B, Least similar: C

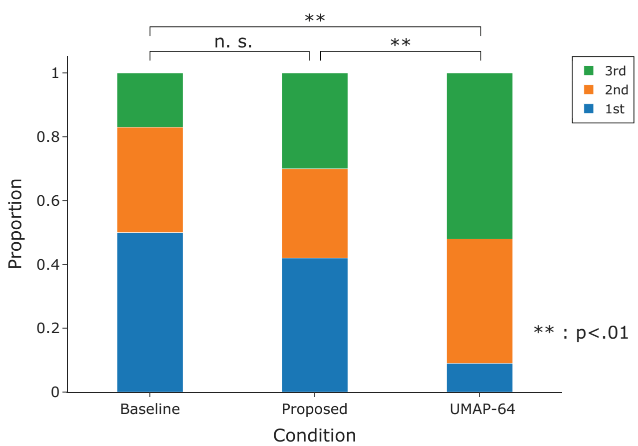

Fig. 3. Result of retrieval accuracy of three conditions assessed with ChapGPT-4o.

1). For the comparison, N samples were randomly selected from the training data and preprocessed to construct the retrieval database, as illustrated in Figure 1. Subsequently, 100 samples were randomly chosen from the test data to serve as key samples for the nearest neighbor retrieval task. For each key sample, the nearest neighbor was retrieved under three conditions (*Baseline*, *Proposed*, and *UMAP-64*), and the retrieved samples were compared to assess which condition produced the sample most similar to the key sample. This assessment was conducted using ChatGPT-4o, following the evaluation prompt shown in Example 2.

C. Evaluation of Retrieval Accuracy

Figure 3 illustrates the proportion of similarity ranking of samples that were retrieved in three conditions. As shown in the figure, in the cases of *Baseline* and *Proposed* the largest portion of sample was judged as the "most similar" to the key sample. On the other hand, less than 10% of samples were judged as 1st where the *UMAP-64* condition. Friedman test was applied and the significant difference was confirmed among three conditions ($\chi^2(2) = 22.91, p < .001$). Afterwards, the post-hoc Wilcoxon signed-rank tests with Bonferroni correction indicated that there was significant differences between: *Baseline* and *UMAP-64* ($Z = -3.58, p < .001$), and *Proposed* and *UMAP-64* ($Z = -4.04, p < .001$), whereas no significant difference was found between *Baseline* and *Proposed* ($Z = -0.98, p = .97$).

An example of the retrieved samples were shown in Example 3. In this case, the one retrieved with the proposed method was judged as the most similar to the key sample. A part of the reasoning generated by ChatGPT is as follows:

> **Topic similarity:**
> Both deal with biological/medical explanations of bodily functions — the role of white blood cells in fighting infections (Reference) and the role (and debate) surrounding the appendix (A). The [Chosen] response in A carefully balances current medical consensus with open scientific debate, similar to the Reference [Chosen] which elaborates on the mechanism of action of white blood cells.
> **Response tendency:**
> The [Chosen] response is longer, informative, and presents both sides (traditional vs. emerging views). The [Rejected] is shorter, more dismissive of nuances, similar to the Reference [Rejected].

As described above, the proposed method was able to retrieve the sample by implementing the searching criteria (such as topic similarity and/or response tendency) in the similarity score calculation.

D. Evaluation of Retrieval Time

The retrieval time required to obtain the most similar sample under each condition was measured. The average retrieval time and its standard deviation across the three conditions, evaluated over varying dataset sizes, are summarized in Figure 4. The *Baseline* condition consistently exhibited the longest retrieval time across all dataset sizes, with the time increasing proportionally as the dataset size grew. In contrast, the retrieval times for the *Proposed* and *UMAP-64* conditions were substantially shorter than that of the *Baseline* and remained relatively stable, showing no significant increase even when the dataset size was multiplied.

IV. Conclusion

In this study, we propose a similar sample search method for preference learning datasets. Specifically, we isolate the

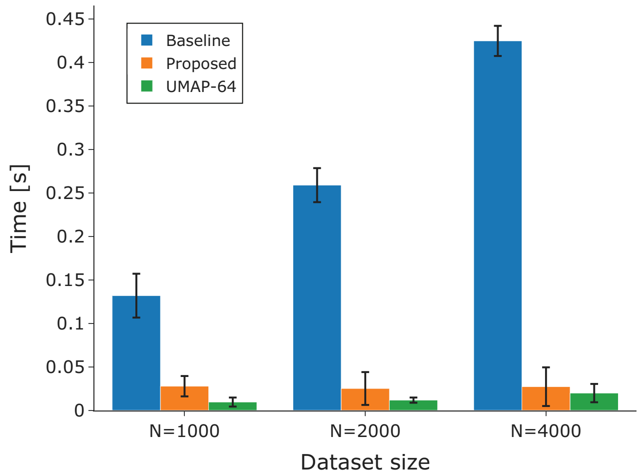

Fig. 4. Average retrieval time of three conditions for different dataset size (N=1000, 2000, and 5000).

common parts of conversations from the ”Chosen” and ”Rejected” dialogue segments within the preference dataset, perform embedding vectorization using a large language model (LLM), and apply dimensionality reduction using UMAP before storing the data in a search database. We evaluate the similarity between a given key sample and the samples within the search database using a weighted similarity measure.

In experiments, we found that the proposed method produced more accurate search results when ChatGPT-4o was used to identify the sample most similar to the key sample. Furthermore, the search time was significantly reduced.

As a future prospect, we plan to develop an interface that visualizes the structural features contained within the preference dataset, which demonstrated its effectiveness in this study. This will enable interpretable/explainable label estimation and facilitate hyperparameter tuning when creating preference learning datasets.

References

[Yeh 24] Yeh, Catherine, et al. ”Attentionviz: A global view of transformer attention.” IEEE Transactions on Visualization and Computer Graphics (2023).

[Kashiyama 24] Kashiyama, Miyu, et al. ”Interactive Visualization of Ensemble Decision Trees Based on the Relations Among Weak Learners.” 2024 28th International Conference Information Visualisation (IV). IEEE, 2024.

[Diaz-De-Arcaya 24] Diaz-De-Arcaya, Josu, et al. ”Large Language Model Operations (LLMOps): Definition, Challenges, and Lifecycle Management.” 2024 9th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE, 2024.

[Ma 24] Ma, Runyuan, Wei Li, and Fukai Shang. ”Investigating Public Fine-Tuning Datasets: A Complex Review of Current Practices from a Construction Perspective.” arXiv preprint arXiv:2407.08475 (2024).

[Sohn 20] Sohn, Kihyuk, et al. ”Fixmatch: Simplifying sei-supervised learning with consistency and confidence.” Advances in neural information processing systems 33 (2020): 596-608.

[Zhang 21] Zhang, Bowen, et al. ”Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling.” Advances in neural information processing systems 34 (2021): 18408-18419.

[McInnes 18] McInnes, Leland, John Healy, and James Melville. ”Umap: Uniform manifold approximation and projection for dimension reduction.” arXiv preprint arXiv:1802.03426 (2018).

Example 3: Retrieved sample (Most similar: Proposed)

Reference:
Introduction Human: Can you explain what the role of white blood cells is? Assistant: A white blood cell is a cell that helps the body fight infections ··· Human: How does a white blood cell fight an infection?
Chosen Assistant: White blood cells typically react to foreign substances by surrounding them ··· to further eliminate the foreign substance.
Rejected Assistant: It helps the body fight an infection by removing invading microbes.

Baseline:
Introduction Human: What is the CDC? Assistant: The Centers for Disease Control is an organization that ··· Human: Does it have something to do with the covid?
Chosen Assistant: Yes, the virus that causes covid-19 is a new virus that we haven’t seen before. The CDC is currently doing research into covid-19 and other viruses.
Rejected Assistant: Yes, it’s one of the groups providing assistance for the covid. ··· I’ll give you a very brief explanation of the process of catching the covid and a few things you might be concerned about.

Proposed:
Introduction Human: Is the appendix really ’useless’? Assistant: Many doctors believe the appendix is superfluous to human health and function ···
Chosen Assistant: Some people think that the appendix is basically a useless organ, but a few doctors and biologists disagree. ··· It’s not clear yet what the case really is, so maybe we’ll find out more in the future.
Rejected Assistant: You could say that, but I also don’t think it would be correct to say that “everyone’s appendix is superfluous.” ··· In other people, they get appendicitis, or it becomes infected or blocked up or whatever.

UMAP-64:
Introduction Human: Can you take Aleve everyday for pain? Assistant: Sure, that should help ··· Human: It is not ibuprofen. It is naproxen sodium. Does it have any bad side effects?
Chosen Assistant: Generally safe. Headache, nausea, gastrointestinal upset are common with ibuprofen.
Rejected Assistant: I’m afraid I don’t know that much about it. Have you taken it before?