1. Background: Learnware Paradigm

Learnware paradigm [Zhou, 2016; Zhou and Tan, 2024]

- Construct a learnware market containing numerous well-performed models and organize them to solve future user tasks by identifying and reusing helpful learnware(s) without building models from scratch.

Learnware components
- Learnware = well-performed model \( \mathbb{A} \) + specification
- Specification describes the specialty and utility of the model.

Procedure of learnware paradigm
- Submitting stage: The learnware market manages submitted models by specifications.
- Displaying stage: The market helps the user identify and reuse helpful learnware(s).

Reduced Kernel Mean Embedding (RKME) specification [Zhou and Tan, 2024]

\[
\min_{\mathbf{w}, \mathbf{k}} \frac{1}{m} \sum_{i=1}^{m} k(\mathbf{z}_i, \mathbf{z}_i) - \sum_{j=1}^{n} \mathbf{w}_j^T k(\mathbf{z}_j, \mathbf{z}_i)
\]

The RKME specification sketches the dataset via weighted samples in RKHS and captures major distribution information without leaking the original data.

2. Contribution

Two specific key issues
- How to characterize model abilities beyond models' original tasks for accurate learnware identification?
- How to avoid examining the entire market for efficient learnware identification?

Evolvable Learnware Specification with Index (ELSI)
- Evolvable specification: Accurate learnware characterization and identification as the market continuously grows.
- Specification index: Organize specifications to ensure efficient operations related to both learnwares and specifications.

3. Evolvable Learnware Specification

Evolvable learnware specification (RKME, \( \mathcal{L}_f \))
- Loss vector \( \mathcal{L}_f \in \mathbb{R}^C \), \( \mathcal{L}_f \) denotes the loss of the model \( f \) on the \( c \)-th RKME: \( \mathcal{L}_f = (\mathcal{L}_f(1), \ldots, \mathcal{L}_f(C)) \).
- The greater information in \( \mathcal{L}_f \) as the market scales up, the better characterization for model \( f \):

\[
\mathcal{L}_f = \sum_{c=1}^{C} \sum_{x \in \mathcal{D}_i} \mathcal{L}_f(x; c)
\]

Challenge for calculating \( \mathcal{L}_f \) specification
- Inefficiency arises from the increasing high dimensions of \( \mathcal{L}_f \) specification as the market continuously grows.

Solution: RKME specification index
- Structurally organize RKMEs via divisive hierarchical clustering to ensure the sparse representation of \( \mathcal{L}_f \) specification.

4. Learnware Identification

The objective for learnware identification
- User data: \( \{x_i\}_{i=1}^{n} \) sampled from \( \mathcal{D}_i \) with the ground-truth function \( h_i \).

\[
f_i = \arg \min_{f \in F} \mathcal{E}_i(f, h_i) = \arg \min_{f \in F} \mathbb{E}_{x \sim D_i} [f(x) - h_i(x)]^2
\]

Two challenges
- Challenge-1: Learnware performance estimation on user task
- Challenge-2: Avoid traversing the market

Solution for learnware performance estimation on user task

\[
\mathbb{E}_{x \sim D_i}[f(x) - h_i(x)]^2 = \mathbb{E}_{x \sim D_i}[f(x) - f(x)]^2 + \mathbb{E}_{x \sim D_i}[f(x) - h_i(x)]^2
\]

Solution for efficient learnware identification
- Using existing hash methods to converting inner product into cosine similarity.

\[
f_i = \arg \min_{f \in F} \mathcal{L}_f(f, h_i) = \arg \min_{f \in F} \mathcal{L}_f(f, h_i)
\]

5. Experiments

Learnware identification performance
- ELSI-traverse achieves the best performance, and ELSI-hash is the efficient version, closely matches it and still outperforms all other contenders.

Learnware identification efficiency
- ELSI-hash achieves the highest efficiency and ELSI-traverse outperforms other contenders in most scenarios.

6. Conclusion

- We make the first attempt to establish evolvable learnware specifications, aiming for increasingly accurate characterization of model abilities beyond their original training tasks as the market continuously grows, thereby constantly facilitating the evolution and enhancement of the overall market capability.
- Through organizing learnwares and constructing specification indexes, we propose an approach called Evolvable Learnware Specification with Index (ELSI), which could achieve evolvable learnware specifications and corresponding efficient learnware identification for users without leaking raw data. As the key components of our approach, specification indexes are established based on the RKME indexed tree and the specification hash table.
- Extensive experimental results on a learnware market encompassing thousands of models and covering six real-world scenarios validate the effectiveness and efficiency of our approach.