ModsNet: Performance-aware Top-k Model Search using Exemplar Dataset

Mengying Wang[+], **Hanchao Ma**[+], Sheng Guan[+], Yiyang Bian[+], Haolai Che[+],
Abhishek Daundkar*, Alp Sehirlioglu*, Yinghui Wu[+]
[+]Depart. of Computer and Data Science, *Depart. of Materials Science and Engineering, CWRU, USA

OVERVIEW

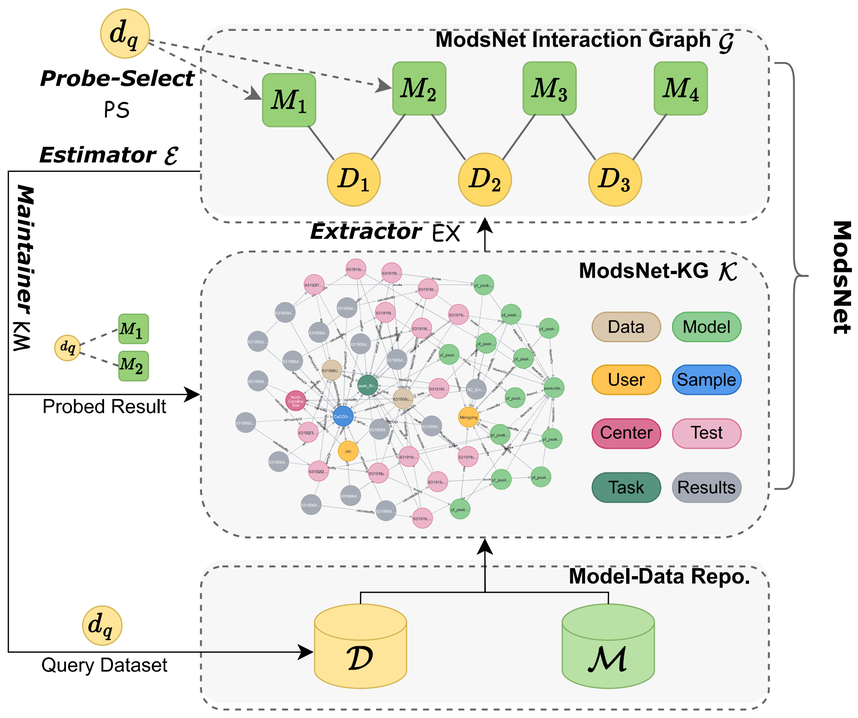
Pre-trained models are invaluable resources, enabling efficient reuse and fine-tuning without starting from scratch. Platforms like HuggingFace gathered abundant models, scripts and associated datasets.

👉 How can we make these models discoverable?

👉 Can we search for models using a 'query' dataset?

ModsNet: a performance-aware data engine designed to search for high-quality, underutilized pre-trained models for data-driven analysis.

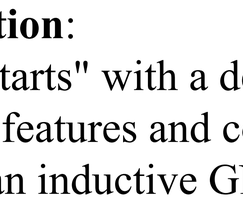
FRAMEWORK



- **Initialization**:
  - "Cold-starts" with a default knowledge graph $\mathcal{K}$.
  - Extract features and construct bipartite graph $\mathcal{G}$.
  - Trains an inductive GNN-based estimator **ε**.
- **Model Selection**:
  - If $d_q \in \mathcal{D}$, and d_q has existing interactions in $\mathcal{G}$, skip probing and use a transductive setting.
  - If $d_q \notin \mathcal{D}$ or no interactions in $\mathcal{G}$, Probe-Selector PS finds promising models, adds probe edges, and consults **ε** for selection.
- **Maintenance**:
  - Maintainer KM logs d_q and updates $\mathcal{K}$ and $\mathcal{G}$.

APPLICATIONS

Query 1:
I have a dataset "tolgadincer/labeled-chest-xray-images", which model should I choose for classifying pneumonia? (k=1)



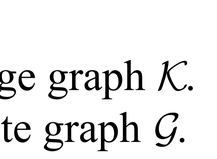
Response 1:
The selected models with estimated balanced accuracy

- Groundtruth{id: 1190, b_accuracy: 0.958}
- ModsNet-C{id: 1175, b_acccuracy: 0.925}
- LinearReg.{id: 1544, b_accuracy: 0.601}

Prediction Result 1:
Prediction result by selected models for the example image in the input dataset

- Groudtruth{pos: 0.9953, neg: 0.0047}
- ModsNet-C{pos: 0.9527, neg: 0.0473}
- LinearReg.{pos: 0.9099, neg: 0.0901}

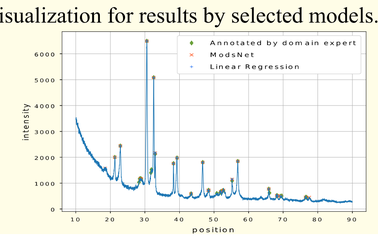
Query 2:
I have a XRDML file, which model should I choose for peak finding? (k=1)



Response 2:
The selected models with estimated f1_score

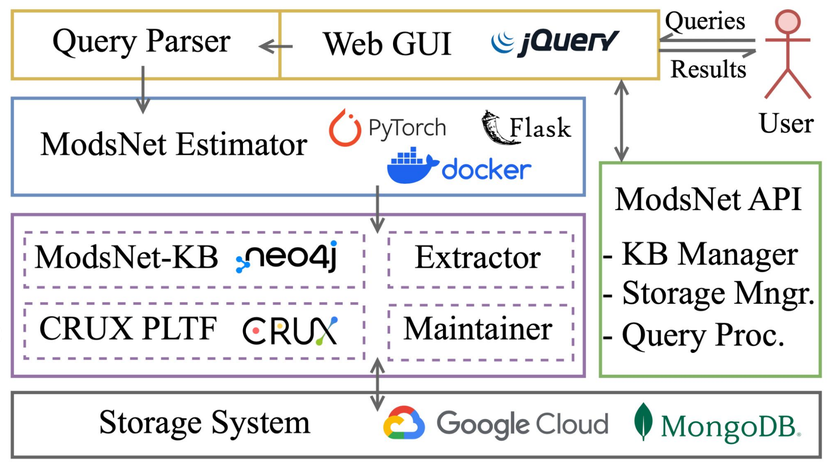
- Groundtruth{id: 311, f1_score: 0.85714}
- ModsNet{id: 291, f1_score: 0.80702}
- LinearReg.{id: 476, f1_score: 0.69388}

Prediction Result 2:
Visualization for results by selected models.



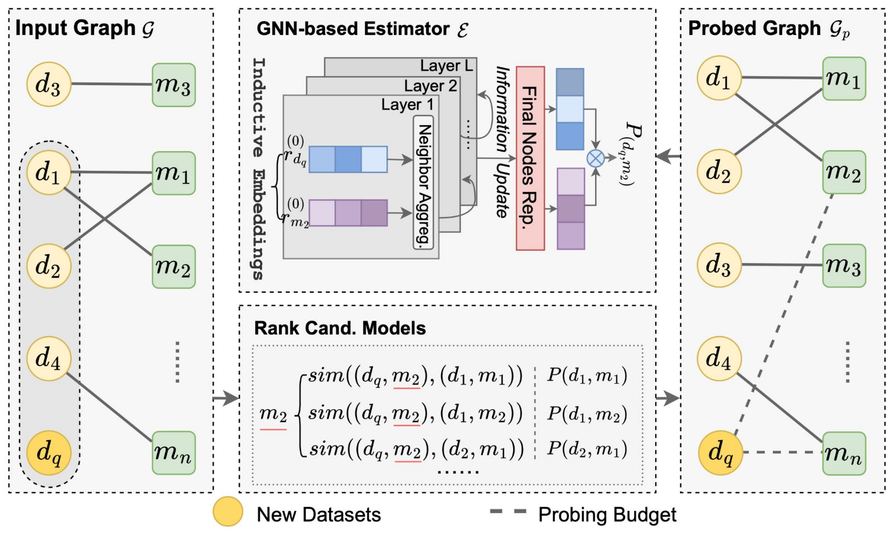
ARCHITECTURE

- **Query Layer**:
  Collects and processes inputs.
- **Recommendation Layer:**
  Estimates performances and selects top-k models by GNN.
- **Knowledge Manag. Layer**:
  Manages the matains knowledge graph and extract features.
- **Storage Layer**: Stores data and models for quick access.



Four-layers Architecture

PROBE-SELECTOR



- **Probe**: Samples promising models and inserts virtual probe edges into $\mathcal{G}$ for verification. Scored by:

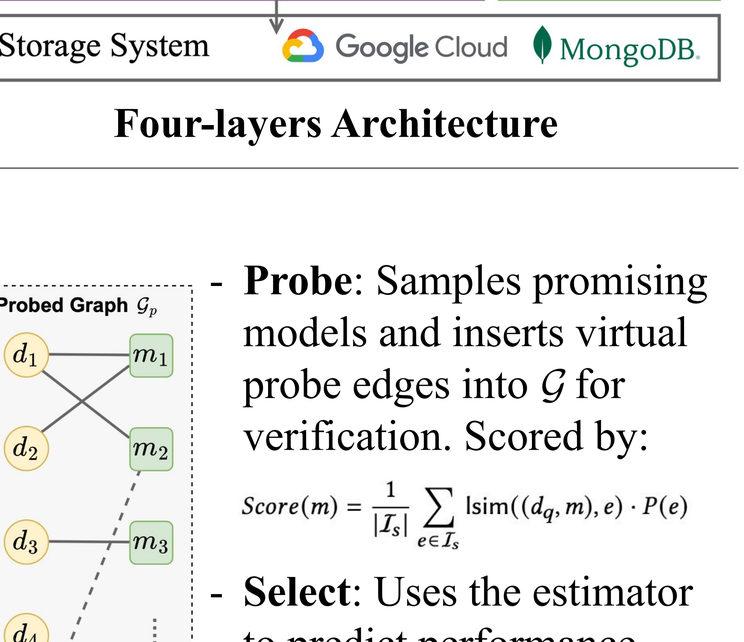
$$Score(m) = \frac{1}{|\mathcal{I}_s|} \sum_{e \in \mathcal{I}_s} \text{lsim}((d_q, m), e) \cdot P(e)$$

- **Select**: Uses the estimator to predict performance, ranks the models, and returns the top-k models.

Learning objective of **ε** that minimizes MSE: $\mathcal{L}_{\text{pred}} = \frac{1}{|\mathcal{I}_{\text{train}}|} \sum_{(d_i,m_j) \in \mathcal{I}_{\text{train}}} (\hat{P}_{(d_i,m_j)} - P_{(d_i,m_j)})^2 + \lambda \|\Theta\|_2^2$

DEMONSTRATION

- **Model Libraries**. Three built-in libraries (KIZoo, HFZoo, PKZoo) with the option to upload new ones.
- **Query by Exemplar Dataset**. Users can find promising top-*k* models without writing complex queries.
- **Performance-aware**. It selects models that outperform baselines in effectiveness and efficiency.
- **Visual Performance Analysis**. Offer tools for interactive evaluation.



BUILD-IN MODEL LIBRARIES

Dataset	# Models	# Datasets	# Interaction	# Features	Density
PKZoo	462	289	98257	21	0.73591
KIZoo	1800	72	9304	41	0.07179
HFZoo	932	66	974	13	0.01583

CRUX
- Peak-finding models, XRD datasets.
- Crowdsourced from material science community, keep growing.

kaggle
- Image datasets are collected from Kaggle.
- Self-curated, over 1,000 GPU hours, various CNN architectures.
- Recorded detailed training and testing information.

Hugging Face
- Text classifiers, text datasets.
- Crowdsourced from a fast-growing AI community.

ACKNOWLEDGEMENT

This work is supported by NSF under OAC-2104007.

ModsNet TEAM

crux-project.github.io