

ModelTable: A Corpus of Tables about Models

Zhengyuan Dong¹, Victor Zhong¹, Renee J. Miller¹
1. Cheriton School of Computer Science, University of Waterloo



Code



Paper

Data Lab

Motivation

Model tables reveal structured semantics beyond text-only retrieval.
Query Table: BERT

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	Average
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35	81	86	61.7	74
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36	73.3	84.9	56.8	71
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80	82.3	56	75.1
BERT_BASE	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT_LARGE	86.7/85.9	72.1	92.7	94.9	60.5	86.5	89.3	70.1	82.1

Table 1: RoBERTa

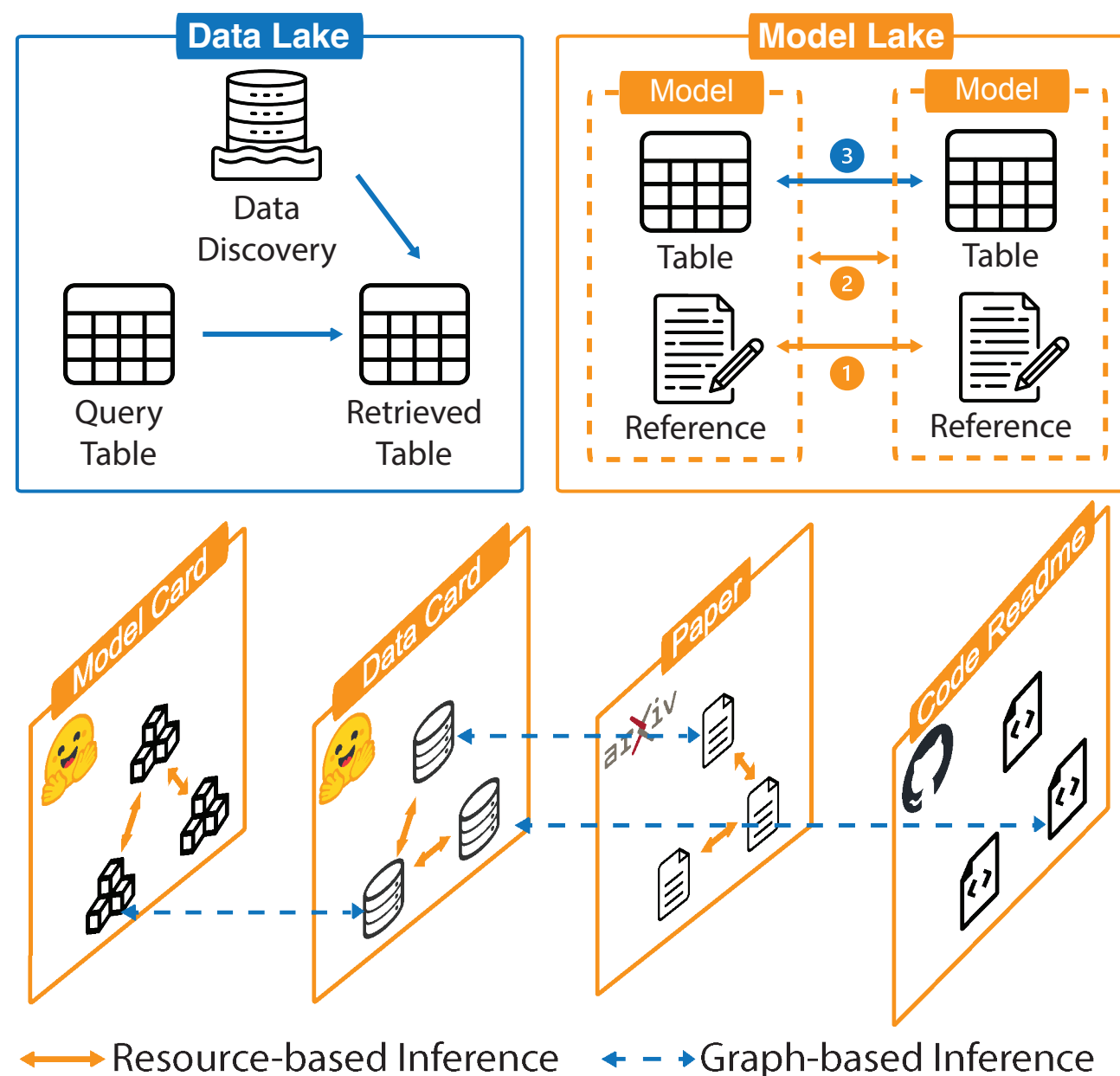
	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
Single-task single models on dev										
BERT_LARGE	86.6/-	92.3	91.3	70.4	93.2	88	60.6	90	-	-
XLNet_LARGE	89.8/-	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa	90.2/90.2	94.7	92.2	86.6	96.4	90.9	68	92.4	91.3	-
Ensembles on test (from leaderboard as of July 25, 2019)										
ALICE	88.2/87.9	95.7	90.7	83.5	95.2	92.6	68.6	91.1	80.8	86.3
MT-DNN	87.9/87.4	96	89.9	86.3	96.5	92.7	68.4	91.1	89	87.6
XLNet	90.2/89.8	98.6	90.3	86.3	96.8	93	67.8	91.6	90.4	88.4
RoBERTa	90.8/90.2	98.9	90.2	88.2	96.7	92.3	67.8	92.2	89	88.5

Table 2: ELECTRA

Model	Train FLOPs	CoLA	SST	MRPC	STS	QQP	MNLI	QNLI	RTE	WNLI	Avg.*	Score
BERT	1.9e20 (0.06x)	60.5	94.9	85.4	86.5	89.3	86.7	92.7	70.1	65.1	79.8	80.5
RoBERTa	3.2e21 (1.02x)	67.8	96.7	89.8	91.9	90.2	90.8	95.4	88.2	89	88.1	88.1
ALBERT	3.1e22 (10x)	69.1	97.1	91.2	92	90.5	91.3	-	89.2	91.8	89	-
XLNet	3.9e21 (1.26x)	70.2	97.1	90.5	92.6	90.4	90.9	-	88.5	92.5	89.1	-
ELECTRA	3.1e21 (1x)	71.7	97.1	90.7	92.5	90.8	91.3	95.8	89.8	92.5	89.5	89.4

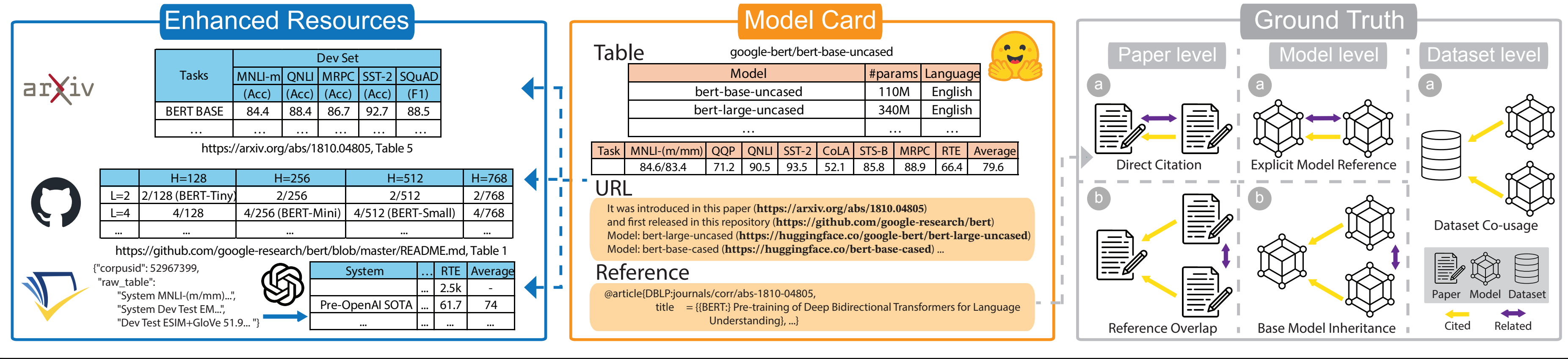
Applications

Supports model discovery and card verification via table relationships.



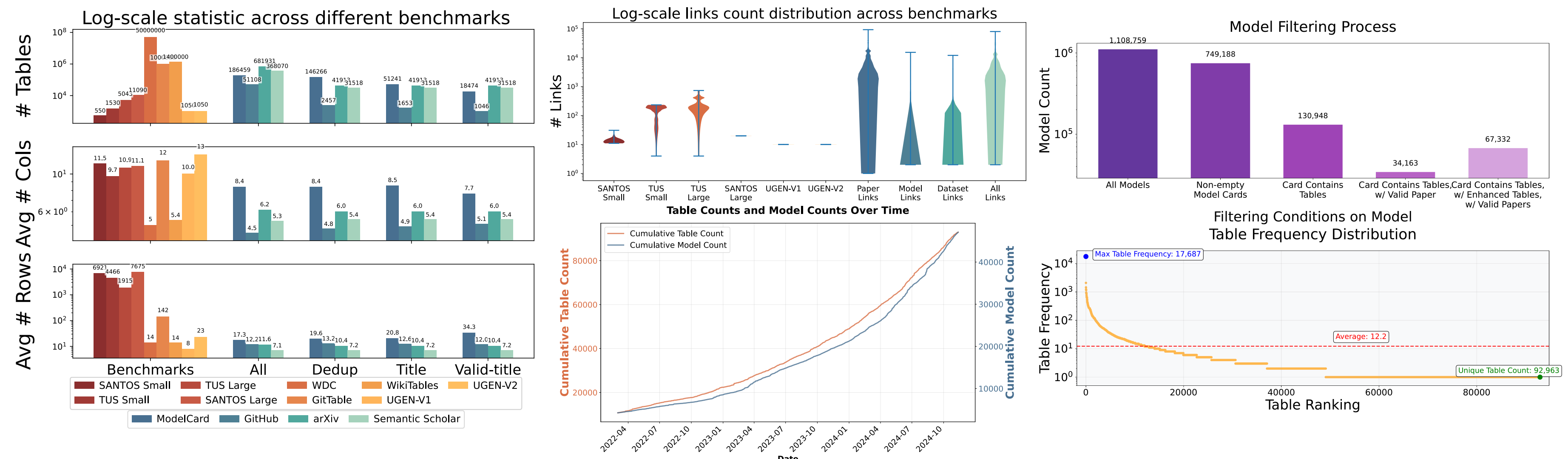
Construction

Build by linking tables from model cards, GitHub, and papers into a unified benchmark with verified model and dataset relations.



Statistics

ModelTable contains over 90K tables across 60K models, revealing denser and more meaningful inter-table links.



Findings

- [1] While All Modern Methods Significantly Outperform Keyword Search, IR methods Outperform Data Lake Methods except Union Search.
- [2] Significant Performance Variation Arises from Differences in Table Source Quality.
- [3] Domain-Specific Augmentations Enhance the Searchability of Tables.
- [4] Robustness and Consistent Performance are Maintained Across Various Ground-Truths.
- [5] Structural augmentations are effective in improving retrieval robustness alongside semantic augmentations.

References

- [1] Koyena Pal, David Bau, and Renée J. Miller. Model Lakes. CoRR abs/2403.02327 (2024).
- [2] Grace Fan, Jin Wang, Yuliang Li, Dan Zhang, and Renée J. Miller. Semantics-aware Dataset Discovery from Data Lakes. PVLDB 2023.
- [3] Mahdi Esmailoghli, Christoph Schnell, Renée J. Miller, and Ziawash Abedjan. BLEND: A Unified Data Discovery System. ICDE 2025.
- [4] Aamod Khatiwada, Roee Shraga, Wolfgang Gatterbauer, and Renée J. Miller. SANTOS: Relationship-based Semantic Table Union Search. SIGMOD 2023.