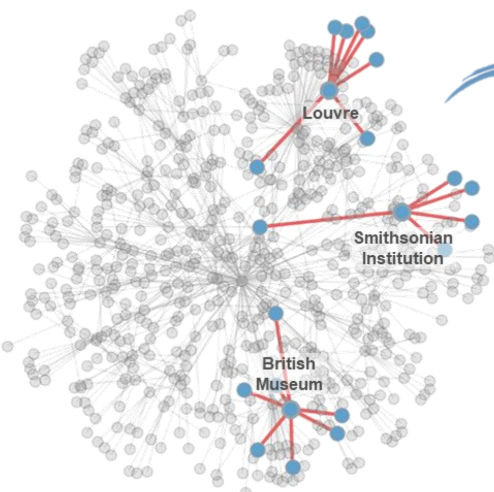
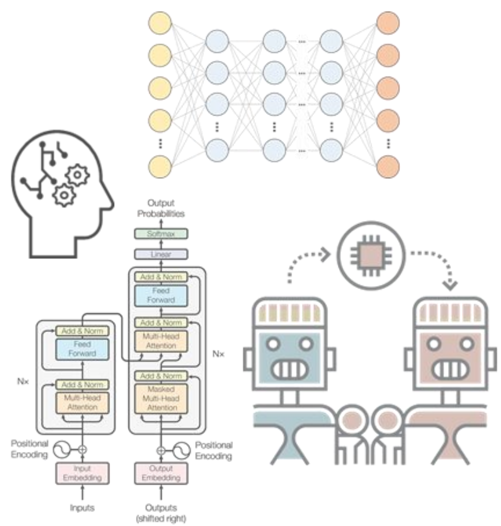
Knowledge pre-training for tabular learning

Jun Kim





Pretrain & Transfer

Outline

Introduction

Knowledge pre-training for tabular learning

CARTE: Context-Aware Representation of Table Entries

TARTE: Transformer Augmented Representation of Table Entries

Empirical studies

Discussion

Introduction

Knowledge pre-training for tabular learning

- CARTE: Context-Aware Representation of Table Entries
- TARTE: Transformer Augmented Representation of Table Entries

Empirical studies

Discussion

Tabular Data

country	descriptior	designatio	points	price	province	region_1	region_2	taster_nar	title	variety	winery
Italy	Aromas in	VulkÃ Biar	87		Sicily & Sa	Etna		Kerin Oâ€	Nicosia 20	White Bler	Nicosia
Portugal	This is ripe	Avidagos	87	15	Douro			Roger Vos	Quinta dos	Portugues	Quinta dos Avidagos
US	Tart and snappy, the f		87	14	Oregon	Willamette	Willamette	Paul Gregu	Rainstorm	Pinot Gris	Rainstorm
US	Pineapple	Reserve La	87	13	Michigan	Lake Michigan Shore		Alexander	St. Julian 2	Riesling	St. Julian
US	Much like	Vintner's R	87	65	Oregon	Willamette	Willamette	Paul Gregu	Sweet Che	Pinot Noir	Sweet Cheeks
Spain	Blackberry	Ars In Vitro	87	15	Northern S	Navarra		Michael Sc	Tandem 20	Tempranill	Tandem
Italy	Here's a br	Belsito	87	16	Sicily & Sa	Vittoria		Kerin Oâ€	Terre di Gi	Frappato	Terre di Giurfo
France	This dry and restraine		87	24	Alsace	Alsace		Roger Vos	Trimbach 2	GewÃ¼rztı	Trimbach
Germany	Savory drie	Shine	87	12	Rheinhessen			Anna Lee (	Heinz Eifel	GewÃ¼rztı	Heinz Eifel
France	This has gr	Les Nature	87	27	Alsace	Alsace		Roger Vos	Jean-Bapti	Pinot Gris	Jean-Baptiste Adam
US	Soft, suppl	Mountain	87	19	California	Napa Valle	Napa	Virginie Bo	Kirkland Si	Cabernet S	Kirkland Signature
France	This is a dry wine, ver		87	30	Alsace	Alsace		Roger Vos	Leon Beye	GewÃ¼rztı	Leon Beyer
US	Slightly reduced, this v		87	34	California	Alexander	Sonoma	Virginie Bo	Louis M. M	Cabernet S	Louis M. Martini
Italy	This is dor	Rosso	87		Sicily & Sa	Etna		Kerin Oâ€	Masseria S	Nerello Ma	Masseria Setteporte
US	Building on 150 years		87	12	California	Central Co	Central Co	Matt Kettr	Mirassou 2	Chardonna	Mirassou
Germany	Zesty oran	Devon	87	24	Mosel			Anna Lee (	Richard BÃ	Riesling	Richard BÃ¶cking
Argentina	Baked plur	Felix	87	30	Other	Cafayate		Michael Sc	Felix Lavac	Malbec	Felix Lavaque
Argentina	Raw black-	Winemake	87	13	Mendoza P	Mendoza		Michael Sc	Gaucho Ar	Malbec	Gaucho Andino
Spain	Desiccatec	Vendimia S	87	28	Northern S	Ribera del Duero		Michael Sc	Pradorey 2	Tempranill	Pradorey

Wine dataset (Wine Enthusiasts)

Abundant tabular data

Tabular data **dominate the data landscape** for enterprises and institutions

The immeasurable **volume** and **central role** in many applications have led numerous **deep learning** methods

The preferred choice, yet remains to be the **Gradient Boosted Decision Trees**

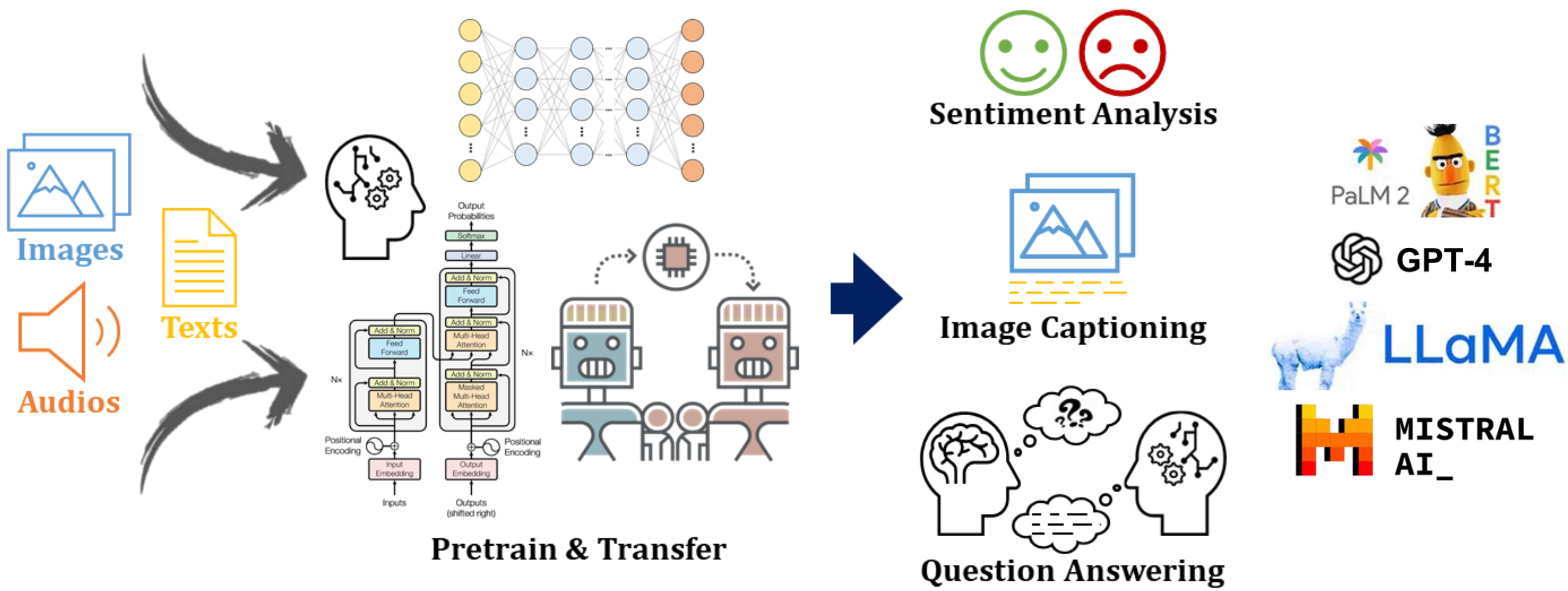
Deep learning to foundation models

The key component of success in **deep learning**: **pretrain and transfer**

Pushing the **next paradigm** to the extreme with **foundation models**

Foundation models:

Any model **trained on broad data** that can be **adapted to a wide range of downstream tasks**

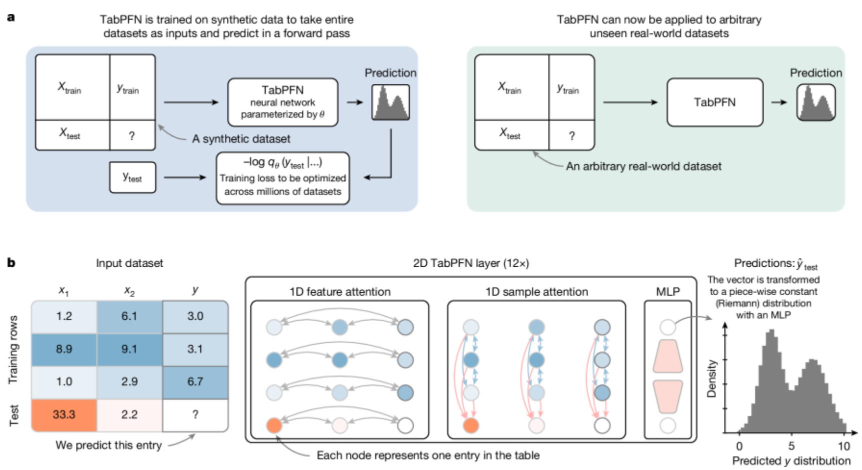


Foundation models for tabular data

Prior Fitted Networks (PFNs)

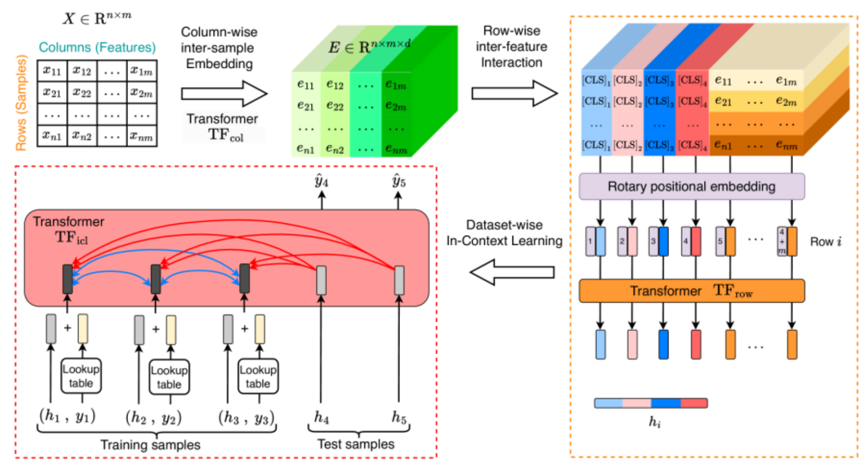
Trained from tremendous amount of synthetic data

High performants for tables of numerical values



TabPFNv2

Hollmann, Noah, et al. "Accurate predictions on small data with a tabular foundation model." Nature 637.8045 (2025): 319-326.



TabICL

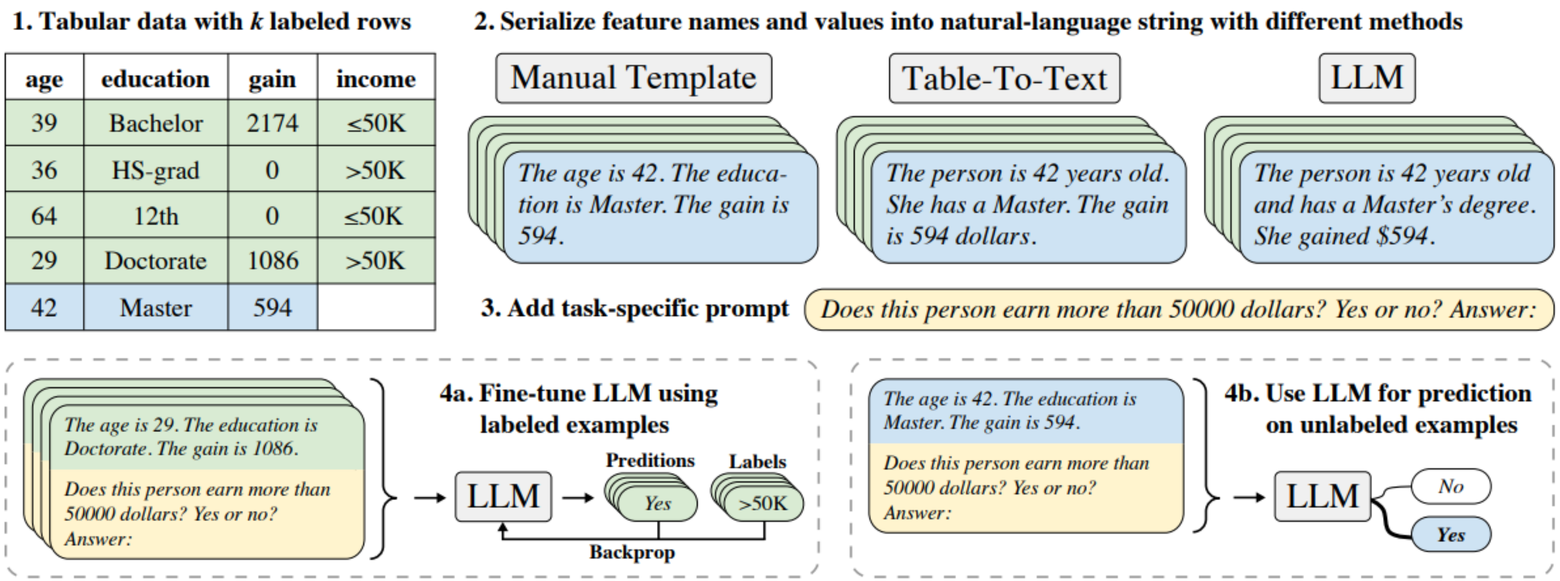
Qu, Jingang, et al. "TabICL: A Tabular Foundation Model for In-Context Learning on Large Data." arXiv preprint arXiv:2502.05564 (2025).

Foundation models for tabular data

Finetune on language models (LLMs)

Serialization of data (creating sentences for each row)

Finetune the LLM on task-specific prompt



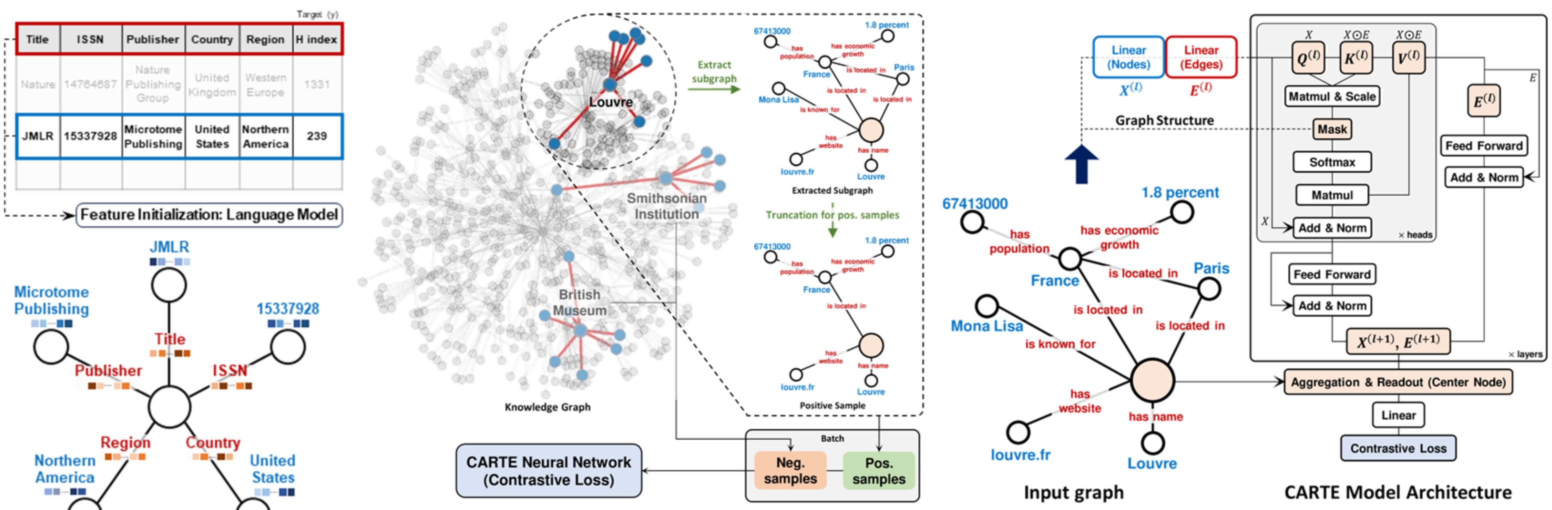
Hegselmann, Stefan, et al. "Tabllm: Few-shot classification of tabular data with large language models." *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023.

Gardner, Josh, Juan Perdomo, and Ludwig Schmidt. "Large scale transfer learning for tabular data via language modeling." *Advances in Neural Information Processing Systems* 37 (2024): 45155-45205.

Foundation models for tabular data

CARTE: Context-Aware Representation of Table Entries

Myung Jun Kim, Léo Grinsztajn, Gaël Varoquaux



[Graph Representation] [Pretrain] [Graph-Attentional Network]

- **Generalization of table entries** with **graph representation**
  - ✓ No schema or entity matching across tables
- **Pretrain** from **large knowledgebases** (knowledge graphs)
  - ✓ Train over 18.1 million triplets of 6.3 million entities in YAGO
- **Graph-attentional network** that **leverage context** in tabular data

Introduction

Knowledge pre-training for tabular learning

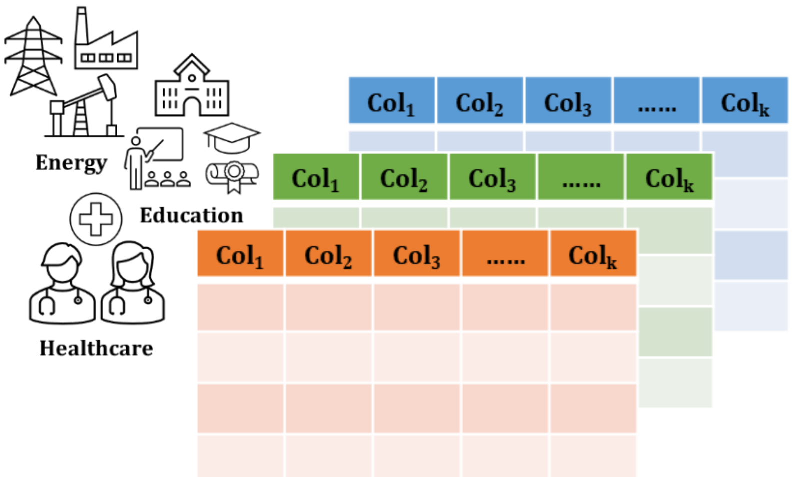
CARTE: Context-Aware Representation of Table Entries

TARTE: Transformer Augmented Representation of Table Entries

Empirical studies

Discussion

Abundant but heterogenous tabular data



Name	Term	Party	VP
George Bush	1989 ~ 1993	Republican	Dan Quayle
Bill Clinton	1993 ~ 2001	Democratic	Al Gore
George Bush	2001 ~ 2009	Republican	Dick Cheney

Name	Incumbency	Party	Country	City
Tony Blair	2007 ~ 2001	Labour	Angleterre	Londres
Nicolas Sarkozy	2001 ~ 2009	Les Républicains	France	Paris

Name	X_1	X_2	...
London	0.0256	0.1267	

Different table specifications

✓ Number of columns, inference tasks

Different entry types

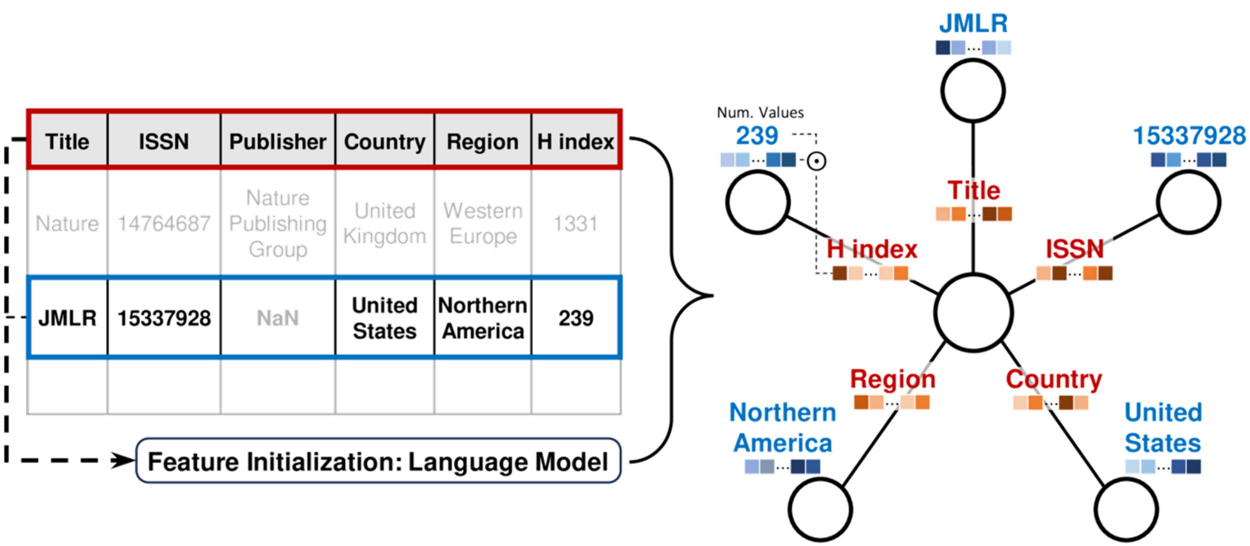
Out of vocabularies

1 "George Bush"? 41st or 43rd president?

2 'Term' and 'Incumbency' are they same columns?

3 'London' and 'Londres ' are they same entities?

Graph representation of table entries



[Graph Representation]

Set the **column name** as the identifier of the **edges**

Set the **value to the corresponding** edge as the **value of the corresponding node**

Use the **language model** to set the **features** for the **nodes** and **edges**

Simple initialization of **numerical values**

The pretrained model: data

YAGO: Yet Another Great Ontology

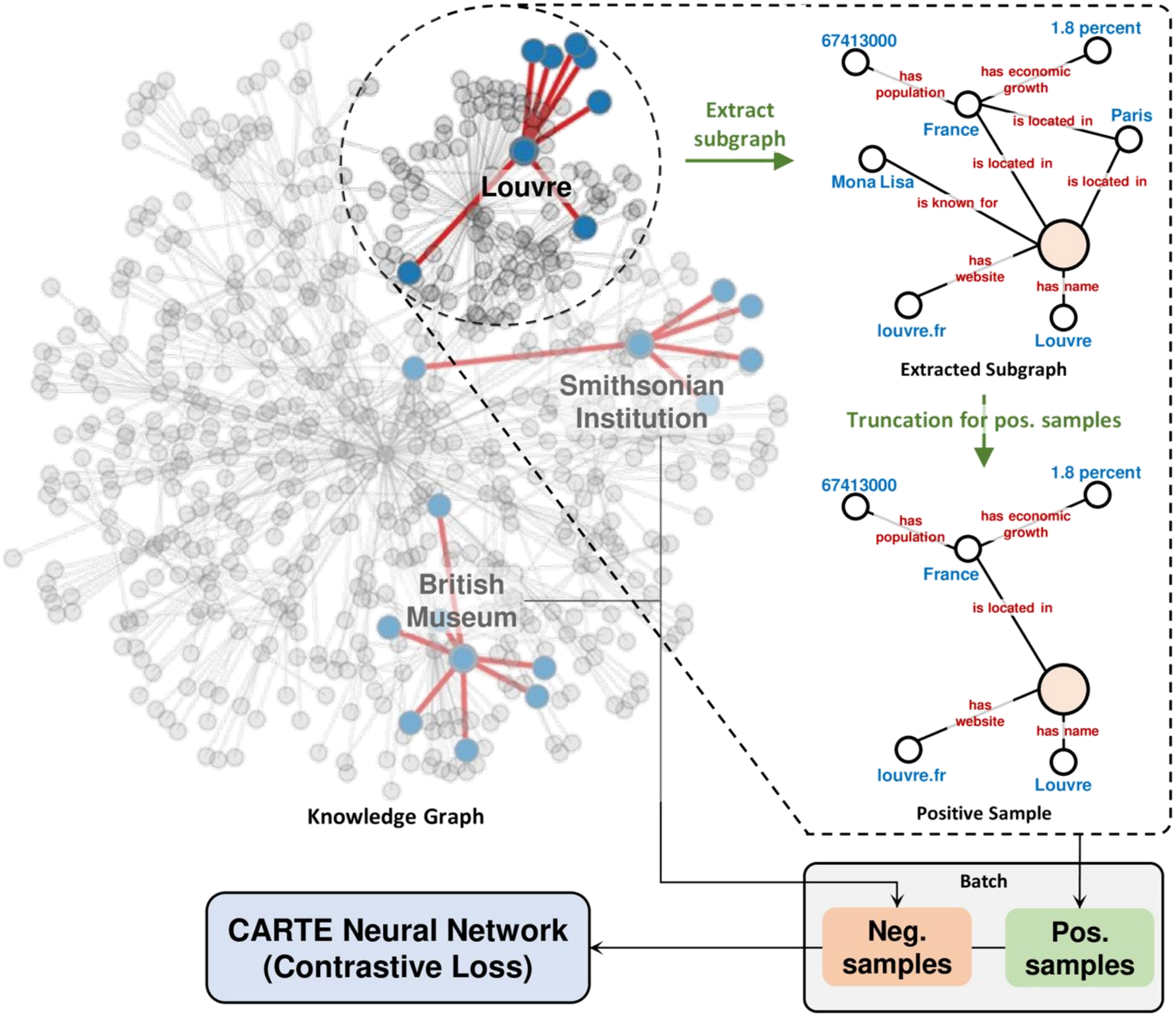
Amarilli, Antoine, et al. "Recent topics of research around the YAGO knowledge base." *Web Technologies and Applications: 16th Asia-Pacific Web Conference, APWeb 2014, Changsha, China, September 5-7, 2014. Proceedings 16*. Springer International Publishing, 2014.

A **large general-purpose knowledge-base** created from Wikipedias with WordNet, GeoNames, etc.,

Contains over **18.1 million triplets** of **6.3 million entities** in YAGO

(Now **5.5 million entities** with **30 million facts**)

CARTE pretraining process



The pretrained model: architecture

Model details

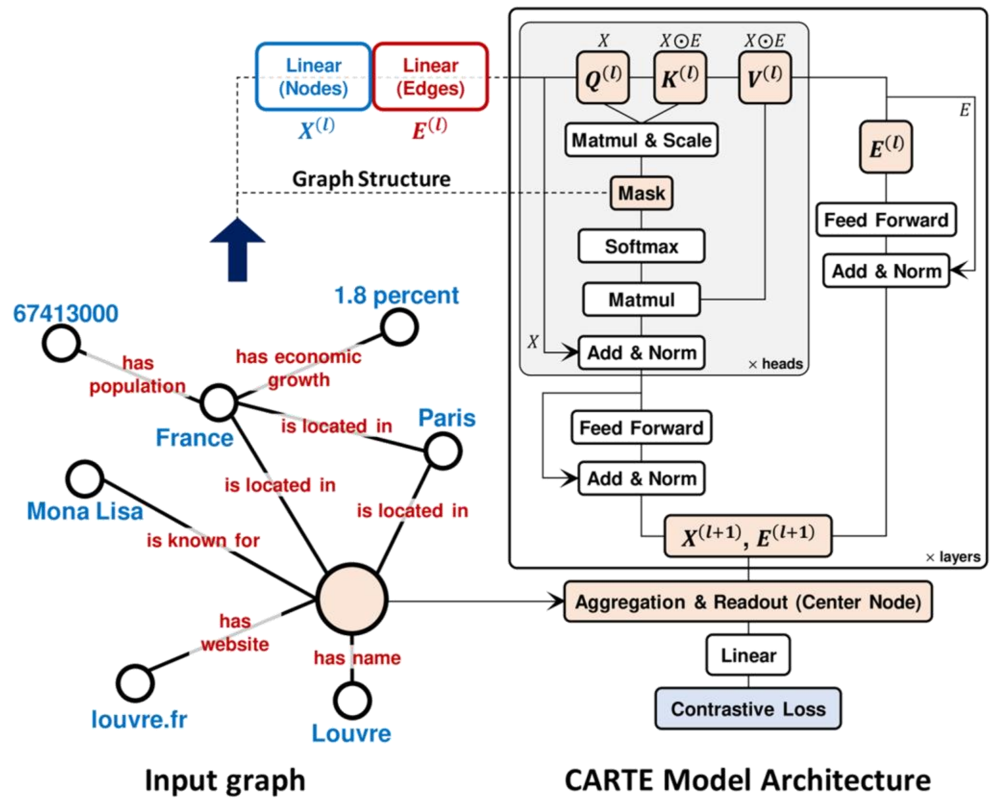
12 layers and multi-head attentions

300 feature dimension

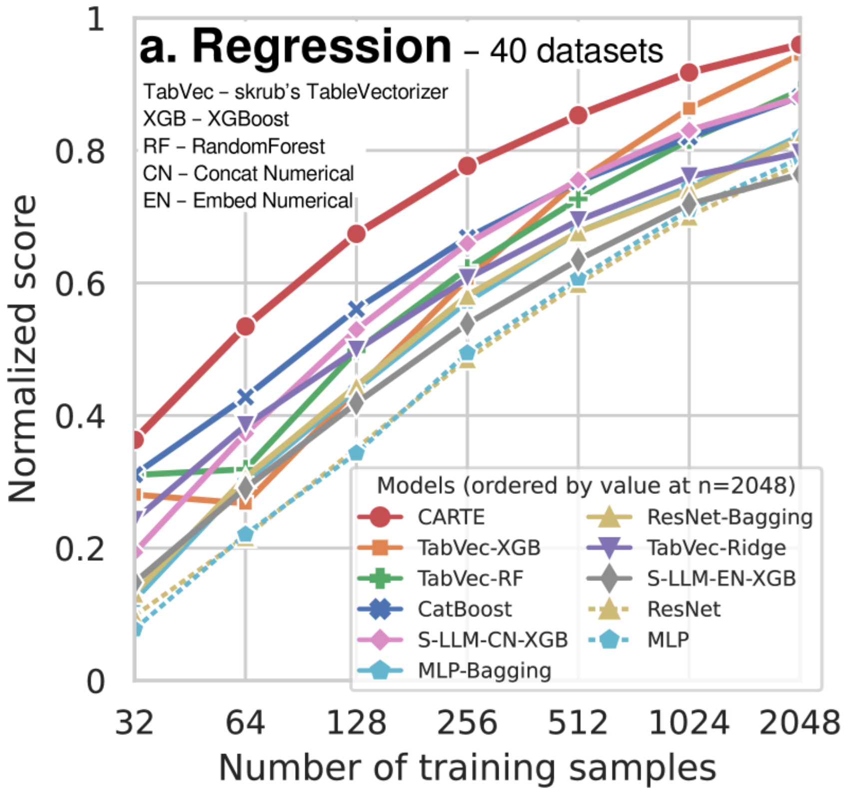
9.3 million parameters

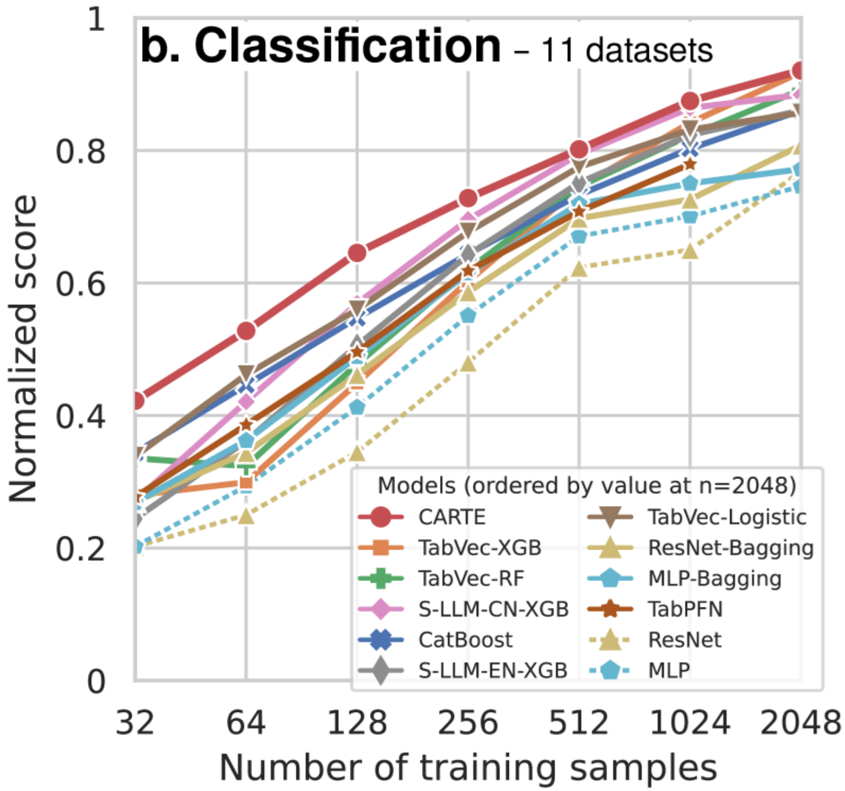
Incorporation of edge information

Use of graph structure



CARTE markedly outperforms the alternatives





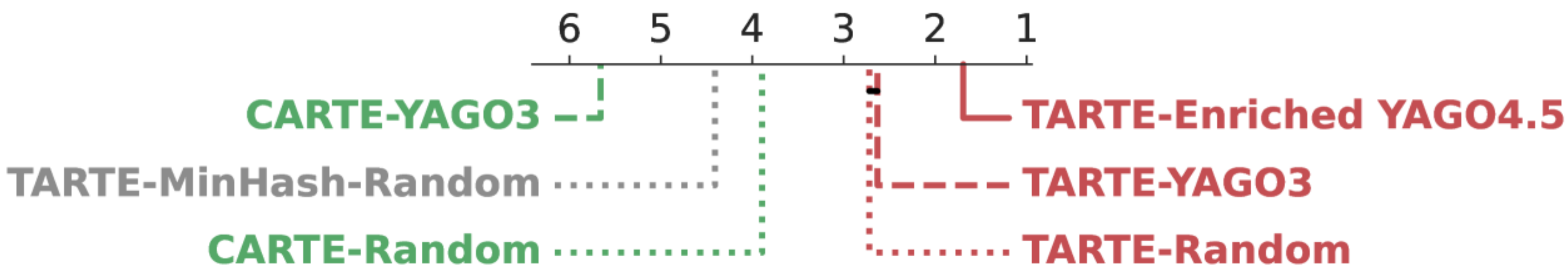
CARTE performs well but…

Computationally expensive

Methods	Preprocessing	Learning $n = 64$	Learning $n = 512$
CARTE	50.20±63.68	85.43±60.30	315.49±119.84
CatBoost	-	0.98±1.19	1.05±1.06
TabVec-XGB	64.72±139.23	0.40±0.21	1.19±0.94
S-LLM-XGB	207.87±361.56	0.87±0.71	3.49±1.79

CARTE performs well but…

Benefits from pretraining ?



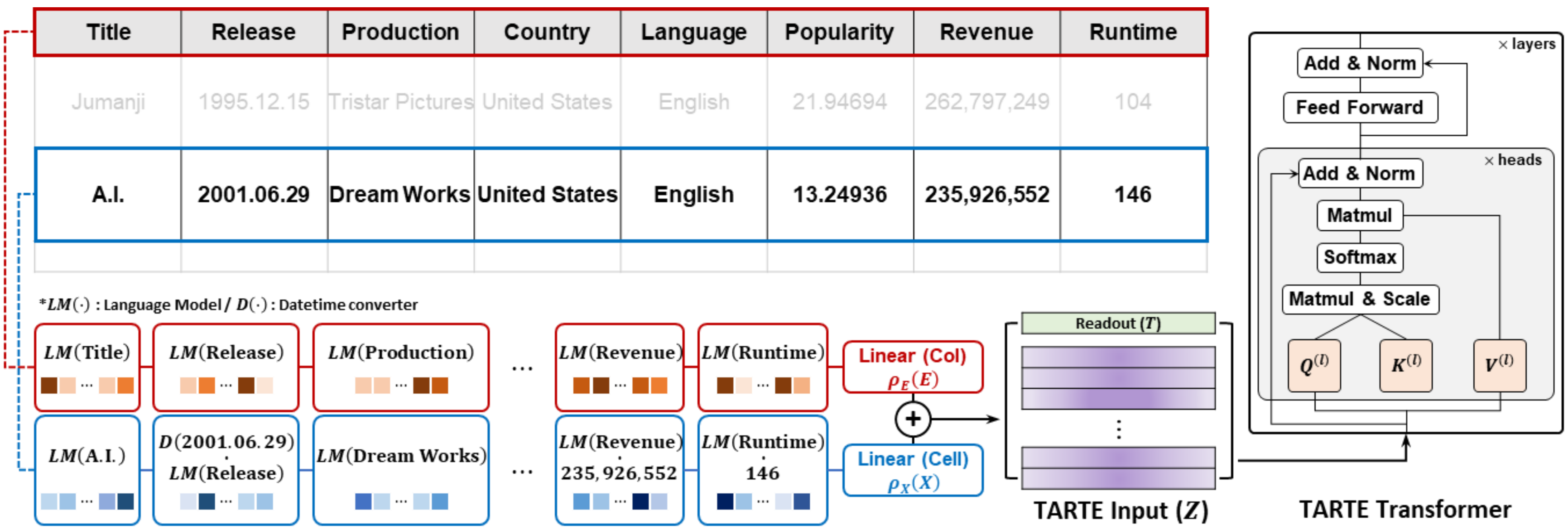
TARTE improves with knowledge pre-training

Transformer with a language model captures well the inductive bias

TARTE – Knowledge pre-training for data semantics

TARTE: Transformer Augmented Representation of Table Entries

Myung Jun Kim, Félix Lefebvre, Gaëtan Brison, Alexandre Perez-Lebel, Gaël Varoquaux



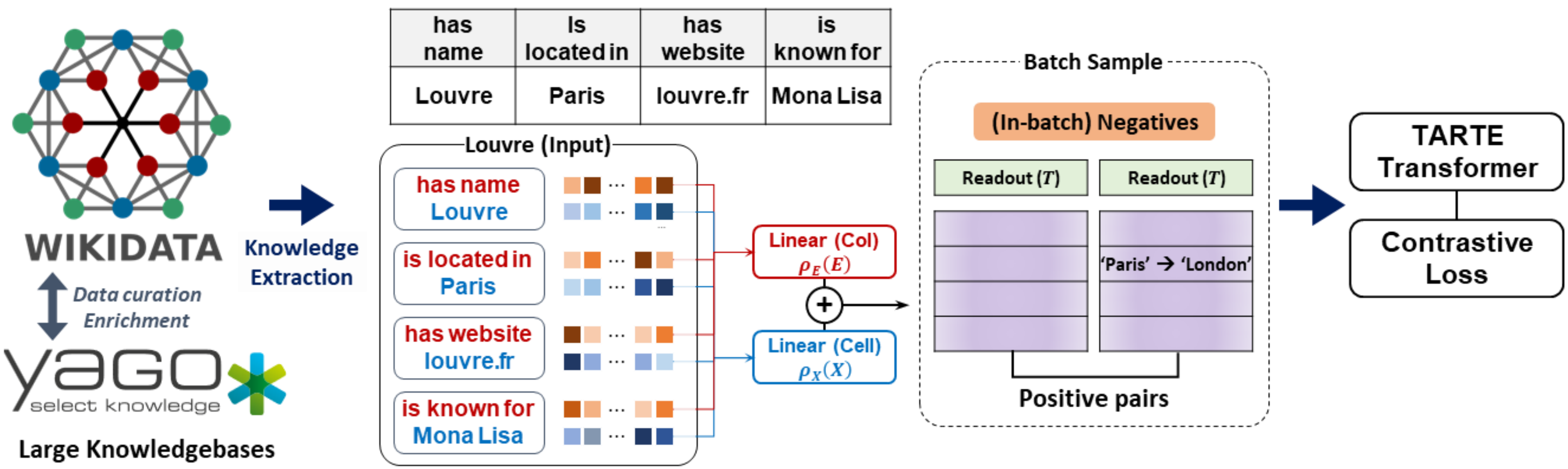
Transformer backbone for knowledge pretraining

More flexible representation of tabular entries

TARTE – Knowledge pre-training for data semantics

TARTE: Transformer Augmented Representation of Table Entries

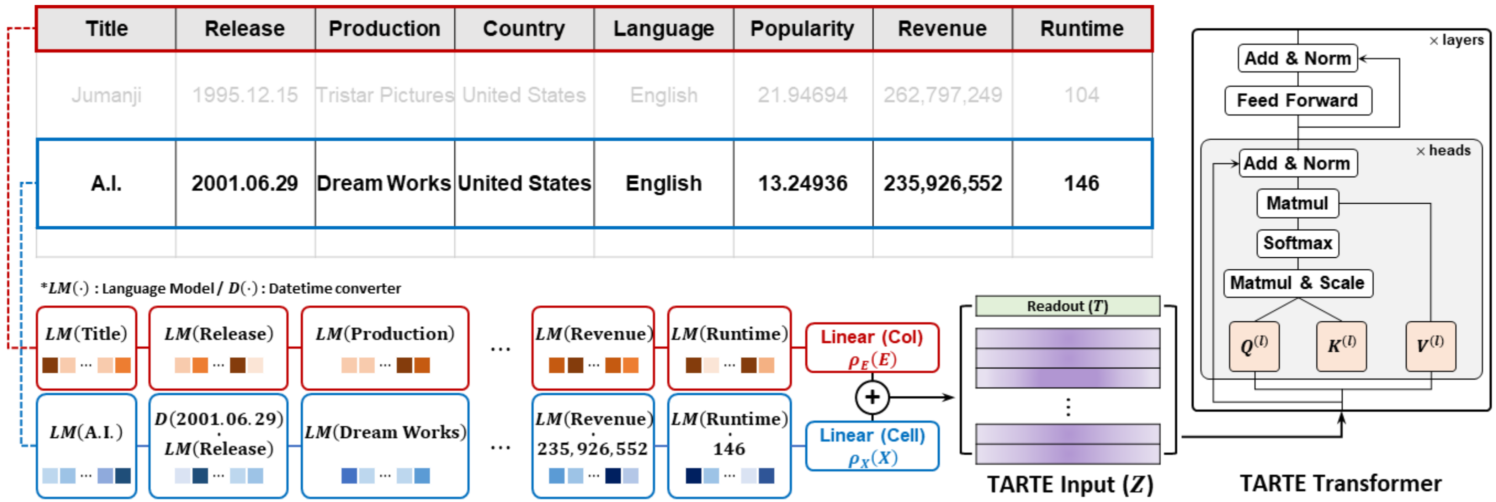
Myung Jun Kim, Félix Lefebvre, Gaëtan Brison, Alexandre Perez-Lebel, Gaël Varoquaux



Larger and **richer** pretrain data to account for heterogeneity

Better control of the pretraining

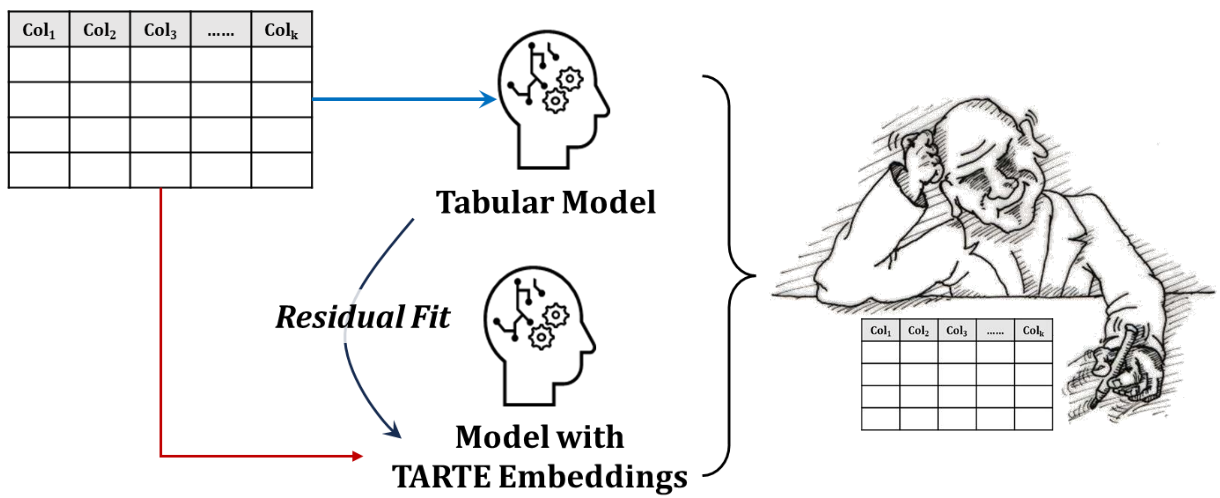
Learning with the backbone



Fine-tuning

TARTE features

TARTE with boosting →

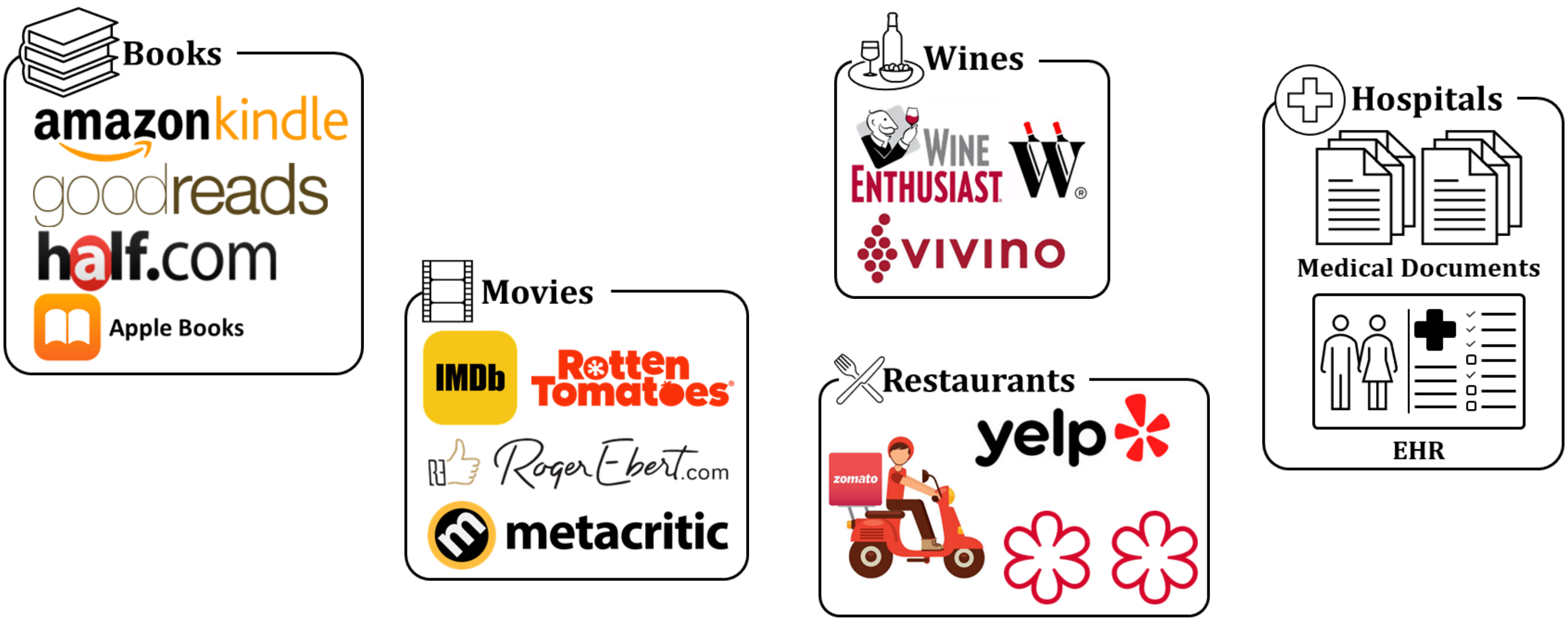


Specializing to a domain

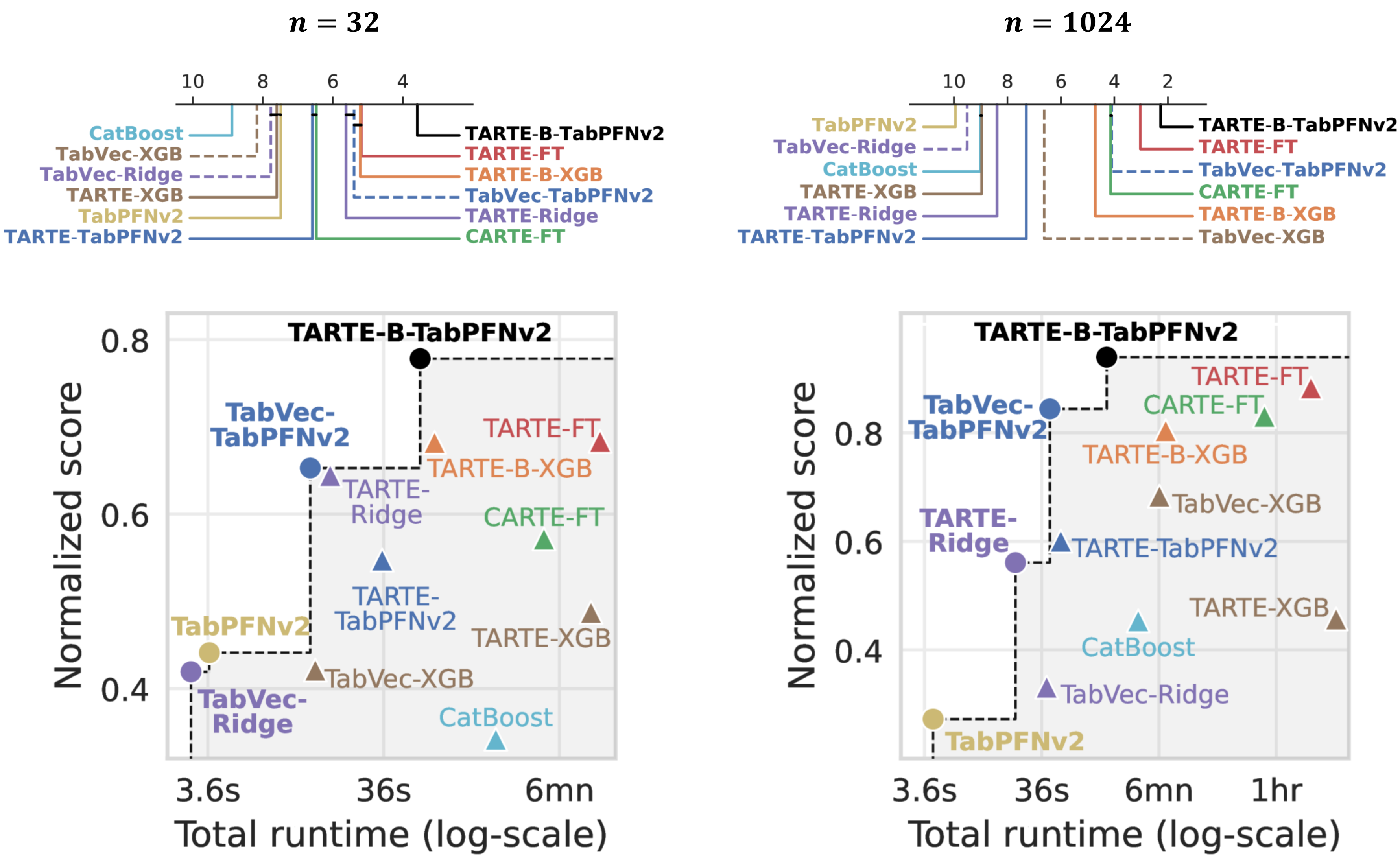
From **fine-tuned models** of related tables, **TARTE** can **readily extract embeddings**

The **TARTE** embeddings can be used with the **boosting scheme**

Residuals are sequentially fitted with embeddings from each fine-tuned model

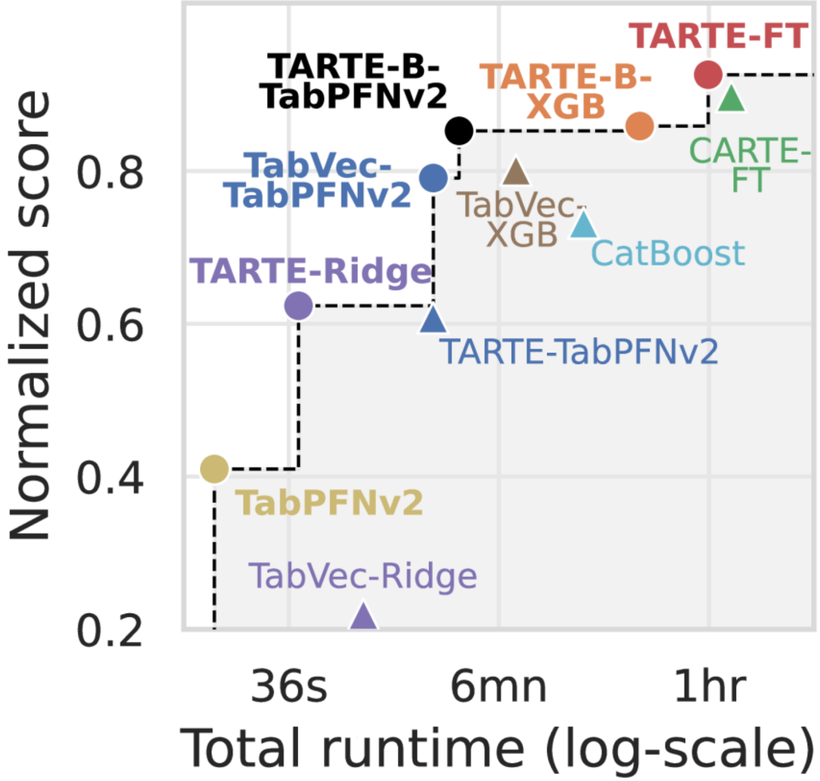


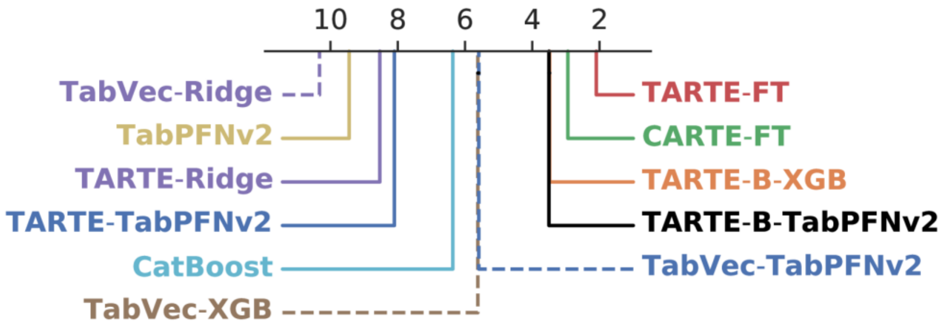
On small tables: few-shot learning



On larger tables (n = 10 000)

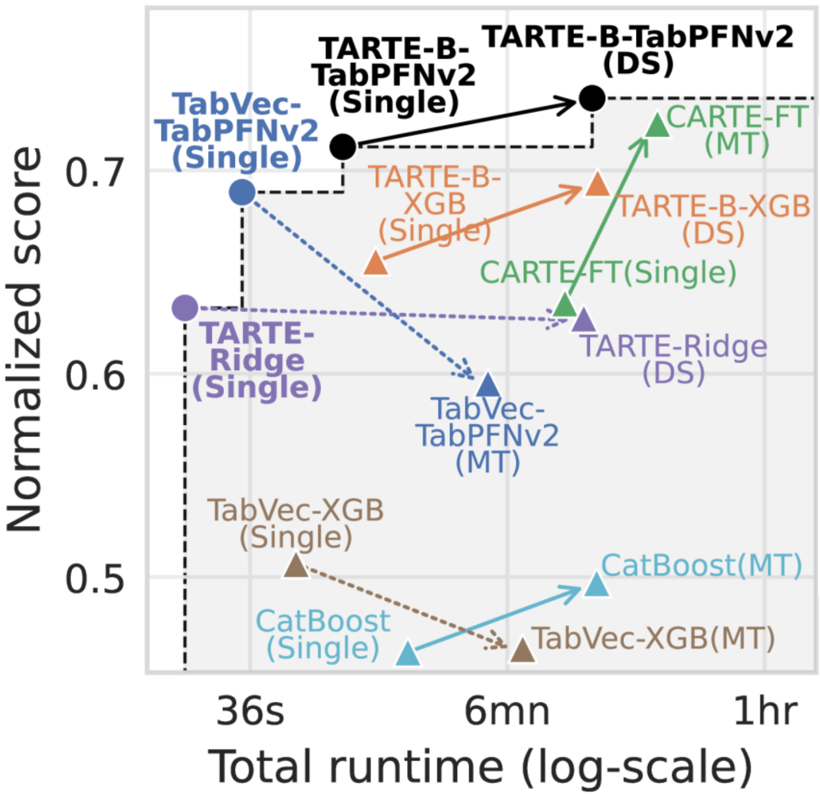
TARTE helps both for prediction and scalability

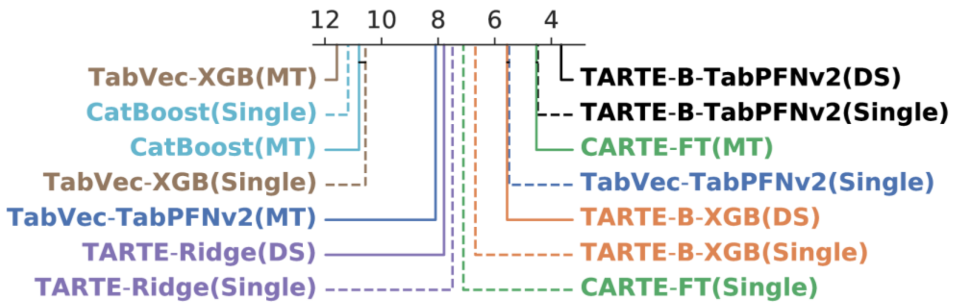




Domain specialization – single source

TARTE improves with **domain specialized representations**





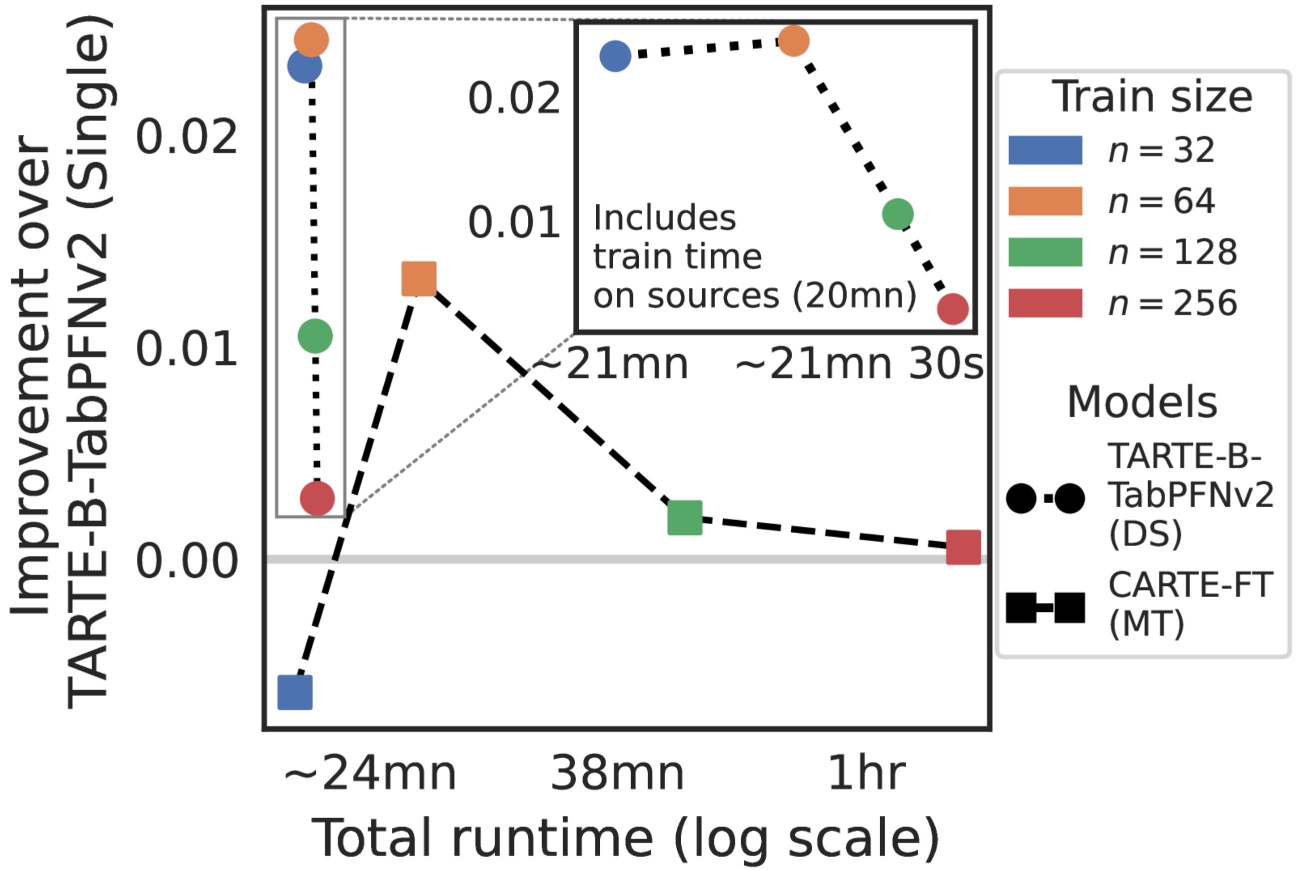
DS: Trained with Domain-Specialized features

MT: Trained on Multi-Tables

Single: Trained on Single table

Domain specialization – multiple sources

TARTE stays efficient with multiple source tables



Introduction

Knowledge pre-training for tabular learning

- CARTE: Context-Aware Representation of Table Entries
- TARTE: Transformer Augmented Representation of Table Entries

Empirical studies

Discussion

TARTE

Knowledge pre-training that helps tabular learning

A re-usable backbone

New research directions for tabular data

Applications to domains specific problems

Thank you