

Relational In-Context Learning on Structured Data via Neighborhood Aggregation and Structural Information (Extended Abstract)

Joe Meyer^{*1}, Tom Palczewski^{*1}, Afreen Shaikh¹, Mahmoud Mohammadi¹, Dinesh Katupputhur Ramprasath¹, Karan Paresh¹, Roshan Upendra Reddy¹, Mark Li¹

¹ SAP Labs, LLC. (Palo Alto), 3410 Hillview Ave., Palo Alto 94304, United States, joseph.meyer@sap.com, tom.palczewski@sap.com

Abstract

Relational databases are one of the most common data storage mechanisms across many business domains including health-care, banking, e-commerce, logistics, and human resources. Traditionally, data science teams must join complex relational database structures into a single table, perform manual feature engineering, and train a single machine learning model, per task. This prevents the exploitation of relational signal and increases the required resources needed to scale across business use cases. There have been significant advancements in foundation model research across computer vision, natural language processing, and tabular deep learning. However, there has been limited work, such as Relational Transformer, exploring models that can directly predict on relational databases. Previous work has applied In-Context Learning (ICL) on tasks that have simpler homogeneous data. We leverage advancements in tabular foundation models (TFM), such as TabPFN and ConTextTab, to directly perform ICL on multi-modal relational data. Specifically, we construct a heterogeneous graph via primary and foreign keys in a relational database. We then apply heterogeneous GraphSAGE model as a fixed random feature map to aggregate neighborhood information across the subgraphs. Additionally, we augment entity node representations by combining structural information with the entity embedding. Finally, we supply a TFM with the contextualized entity node representation to perform ICL on arbitrary tasks - without any training. Our method is tested on the public relational deep learning benchmark Relbench, which contains many diverse, real-world predictive regression and classification tasks across healthcare, e-commerce, marketing, and enterprise sales. In a fully ICL regime, our model is shown to be competitive with several fully-trained benchmarks on classification tasks. On average, the method achieves 114% of the fully-trained LightGBM, 95% of the fully-trained relational deep learning model reported in Relbench, and 98% of the zero-shot performance of pretrained RT. This signifies an advancement in unlocking accurate predictions, without training, on the dominant business data structure across diverse domains.

Supporting Materials

Methodology We randomly initialize a heterogeneous GraphSAGE model (Hamilton, Ying, and Leskovec 2017)

^{*}These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

provided by the RDL pipeline in Relbench (Robinson et al. 2024) pipeline.

$$\mathbf{h}_k = \text{HGS}_{\theta_0}(\mathcal{G}_k), \mathbf{z} = \text{concat}_{k=1}^K \mathbf{h}_k, \hat{\mathbf{y}}_{\text{test}} = \text{TFM}(\mathbf{z}_{\text{train}}, \mathbf{y}_{\text{train}}, \mathbf{z}_{\text{test}}). \quad (1)$$

Here, \mathcal{G}_k denotes the k -th subgraph, HGS_{θ_0} is the randomly initialized encoder. GraphSAGE across relational neighborhoods. Then, we extract each entity node representation and concatenate them with node degree, mean, median, and mode of relative time difference from neighbor nodes, forming \mathbf{z} . Finally, we input the embedded representation \mathbf{z} into TabPFN 2.5 (Grinsztajn et al. 2026), along with the train labels.

Results Table 1 summarizes performance across all datasets and tasks. Our method achieves competitive results while not conducting relational database-specific training or pretraining.

Dataset	Task	LGBM	RDL	RT	Ours	Ours/1	Ours/2	Ours/3
Amazon	I.Churn	62.54	82.81	73.30	77.89	125%	94%	106%
Amazon	U.Churn	52.22	70.42	66.10	64.30	123%	91%	97%
Avito	U.Visits	53.05	66.20	62.60	65.35	123%	99%	104%
Avito	U.Clicks	53.60	65.90	60.90	60.68	113%	92%	100%
Event	U.Repeat	68.04	76.89	-	73.58	108%	96%	-
Event	User.Ignore	79.93	81.62	-	84.13	105%	103%	-
F1	DNF	68.56	72.62	81.20	64.26	94%	88%	79%
F1	Top-3	73.92	75.54	89.30	74.67	101%	99%	84%
HM	U.Churn	55.21	69.88	63.30	68.12	123%	97%	108%
Stack	U.Badge	63.43	88.86	81.10	85.34	135%	96%	105%
Stack	U.Engage	63.39	90.59	89.40	87.48	138%	97%	98%
Trial	Study	70.09	68.60	54.60	56.68	81%	83%	104%
Average	-					114%	95%	98%

Table 1: AUC on RelBench. U = "user", I = "item". Ours/1 (Ours/LGBM), Ours/2 (Ours/RDL), Ours/3 (Ours/RT).

References

Grinsztajn, L.; Flöge, K.; Key, O.; Birkel, F.; Jund, P.; Roof, B.; Jäger, B.; Safaric, D.; Alessi, S.; Hayler, A.; Manium, M.; Yu, R.; Jablonski, F.; Hoo, S. B.; Garg, A.; Robertson, J.; Bühler, M.; Moroshan, V.; Purucker, L.; Cornu, C.; Wehrhahn, L. C.; Bonetto, A.; Schölkopf, B.; Gambhir, S.; Hollmann, N.; and Hutter, F. 2026. TabPFN-2.5: Advancing the State of the Art in Tabular Foundation

Models. ArXiv:2511.08667v2 (last revised 5 Feb 2026), arXiv:2511.08667.

Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. arXiv:1706.02216.

Robinson, J.; Ranjan, R.; Hu, W.; Huang, K.; Han, J.; Dobles, A.; Fey, M.; Lenssen, J. E.; Yuan, Y.; Zhang, Z.; He, X.; and Leskovec, J. 2024. RELBENCH: A Benchmark for Deep Learning on Relational Databases. ArXiv:2407.20060v1 (29 Jul 2024), arXiv:2407.20060.