ECML-PKDD23' ML4ITS WORKSHOP – SEPTEMBER 22, 2023

Probabilistic Demand Forecasting with Graph Neural Networks

Nikita Kozodoi[1]
Elizaveta Zinovyeva[1]
Simon Valentin[2]
João Pereira[3]
Rodrigo Agundez[3]

[1]Amazon Web Services [2]University of Edinburgh [3]adidas AG

Project Team

Amsterdam

Berlin

1. Motivation & Related Work

Demand Forecasting in e-Commerce

Enables optimizing stock planning, logistics, and supply chain operations.

- Ensure product availability online ↔️ Under-prediction
- Minimize waste ↔️ Over-prediction

Demand Forecasting: Related Work

TRADITIONAL TIME SERIES MODELS

ARIMA, moving average, univariate time series models (e.g., [3]).

LIMITATIONS

Cold starts

Scalability

Demand Forecasting: Related Work

TRADITIONAL TIME SERIES MODELS

ARIMA, moving average, univariate time series models (e.g., [3]).

NEURAL SEQUENCE MODELS

RNNs, DeepAR [4], Seq2Seq, Transformer models.

LIMITATIONS

Cold starts

Scalability

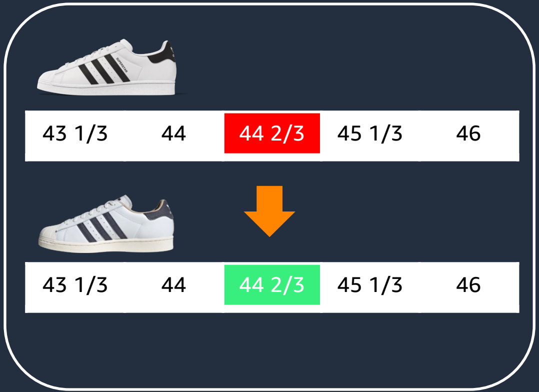
Independent demand predictions

Articles are not “aware” of each other’s existence

Motivation

Article's demand depends on the demand of **related** articles.

Price of similar articles



Stockouts

Demand Forecasting: Related Work

TRADITIONAL TIME SERIES MODELS

ARIMA, moving average, univariate time series models (e.g., [3]).

NEURAL SEQUENCE MODELS

RNNs, DeepAR [4], Seq2Seq, Transformer models.

GRAPH NEURAL NETWORKS

Spatio-temporal GNNs [2].

LIMITATIONS

Cold starts

Scalability

Independent demand predictions

Articles are not “aware” of each other’s existence

Graph NNs in (Demand) Forecasting

- Domains with predefined graph structures
  - Traffic forecasting applications [5]
  - Molecular structures [6]
  - Prominent architectures: DCRNN, Spatio-Temporal GCN, GraphWaveNet
- Demand forecasting in e-Commerce
  - Literature remains **rather limited**
    - Key challenges: high dimensionality, no pre-defined graph structure
  - Previous work combined GNN & LSTM for forecasting in online marketplaces [2]
    - Limitations: multiple-seller setting, point-based forecasts

Our Contributions

- End-to-end forecasting system
  - Based on DeepAR – SOTA LSTM-based forecasting method [4]
  - Integrates Graph-based GNN encoder to account for article relationships
  - Enables probabilistic forecasting
- Generic graph construction approach
  - Does not require expert knowledge and uses data-driven approach
  - Based on article attribute similarity
  - Highly scalable

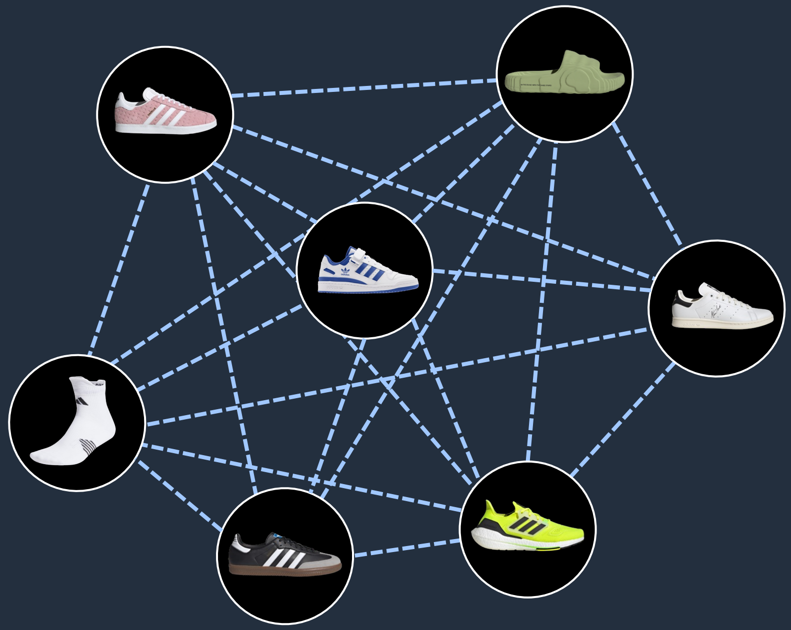
2. Methodology

Graph Construction

Graph Construction [1/3]

- Build a graph based on article similarity
  - Each node represents an article
  - Connections based on cosine similarity
  - Attributes: size, color, category, etc.

$$\text{similarity}\left(\boldsymbol{A}_i, \boldsymbol{A}_j\right) := \cos\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) = \frac{X_i \cdot X_j}{||X_i||||X_j||}$$

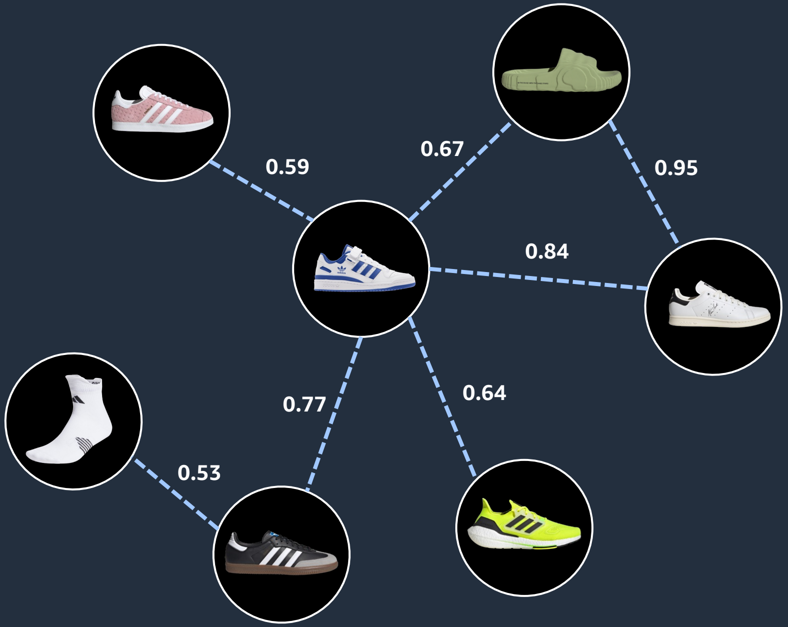


aws

Graph Construction [2/3]

- Build a graph based on article similarity
  - Each node represents an article
  - Connections based on cosine similarity
  - Attributes: size, color, category, etc.
  - Keep edges with similarity > cutoff

$$\text{similarity}\left(\boldsymbol{A}_i, \boldsymbol{A}_j\right) := \cos\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) = \frac{X_i \cdot X_j}{||X_i||||X_j||}$$



Note: similarity numbers are artificial

Graph Construction [3/3]

- Build a graph based on article similarity
  - Each node represents an article
  - Connections based on cosine similarity
  - Attributes: size, color, category, etc.
  - Keep edges with similarity > cutoff
- Nodes include article features
  - Static article attributes
  - Dynamic demand lags

ARTICLE ATTRIBUTES AND DEMAND LAGS

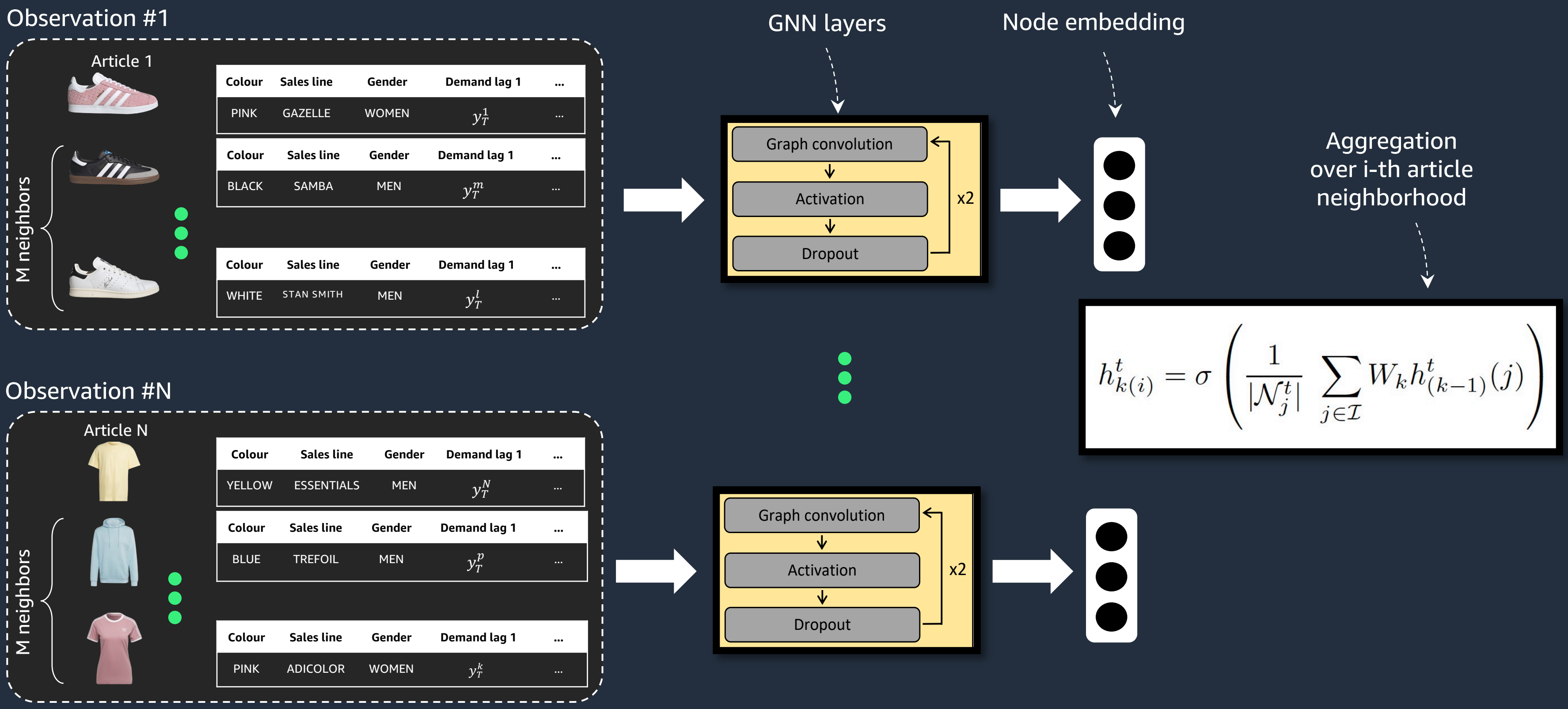
Color	Age	Demand lag 1	...	Demand lag P	Target demand
PINK	ADULT	y_T^i	...	y_{T-P-1}^i	y_{T+1}^i

Note: similarity numbers are artificial

2. Methodology

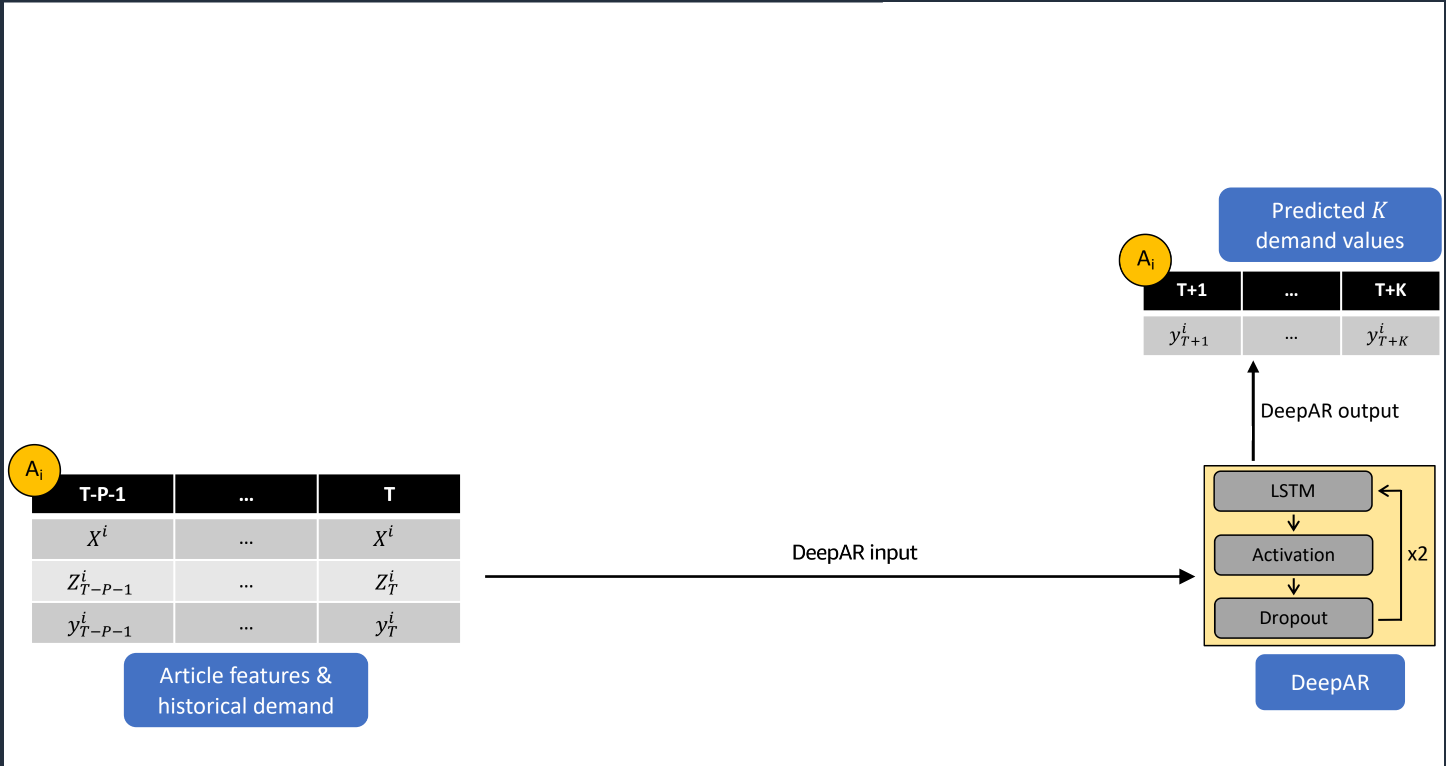
Model Architecture

GNN Encoder

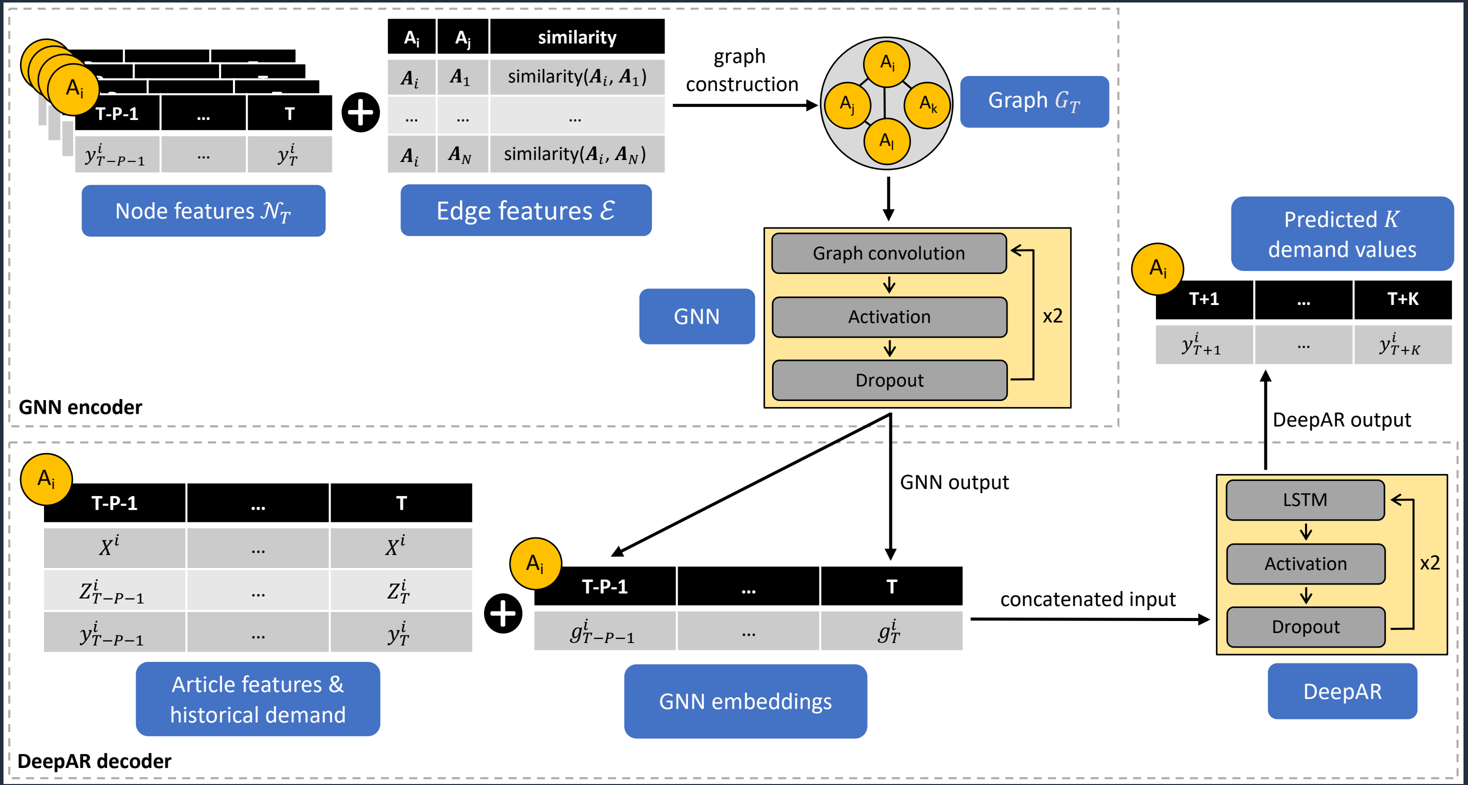


$$h^t_{k(i)} = \sigma\left(\frac{1}{|\mathcal{N}_j^t|}\sum_{j\in\mathcal{I}} W_k h^t_{(k-1)}(j)\right)$$

Model Architecture: Vanilla DeepAR



Model Architecture: GraphDeepAR



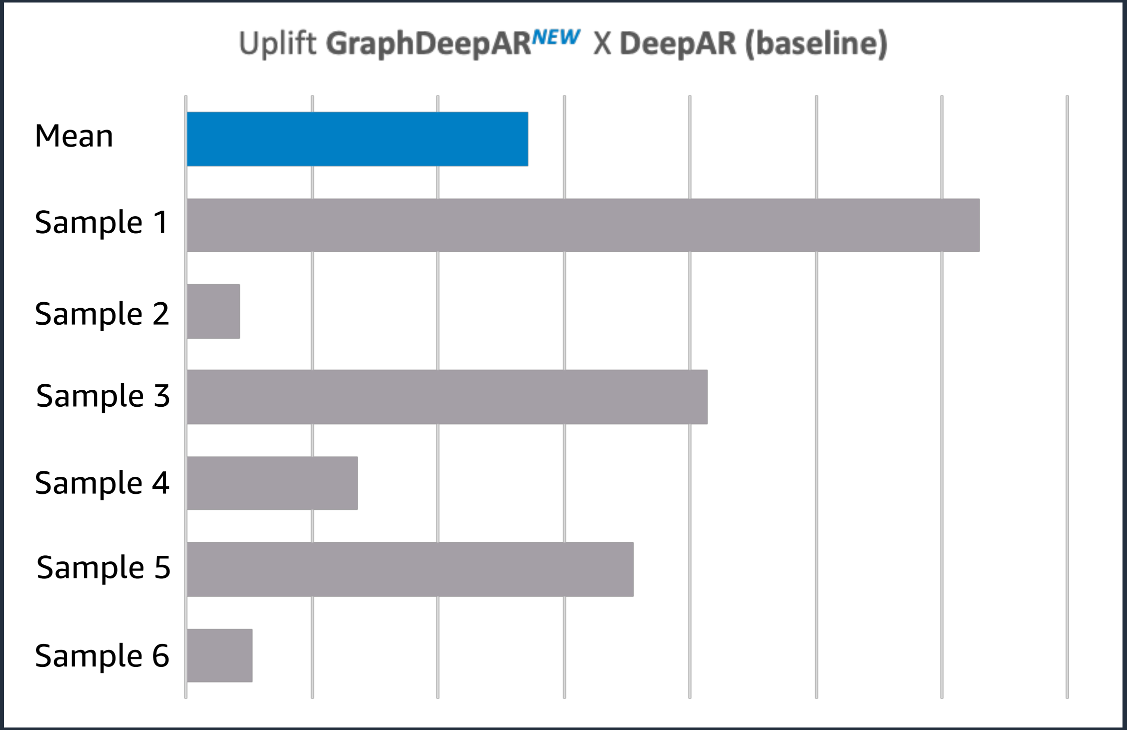
3. Experimental Results

Data Summary

- Two public datasets & one proprietary adidas dataset
- Contain time series with:
  - Article demand
  - Static features (e.g., color, size)
  - Time-varying features (e.g., week number, month number)

Dataset	No. articles	No. weeks	No. features
Retail	629	148	12
E-commerce	8,810	128	5
adidas	80,838	140	20

Performance on adidas Data



Comparing two models:

- **DeepAR (benchmark)**
- **GraphDeepAR (ours)**

GraphDeepAR wins:

- **6/6 times**

Mean financial uplift:

- **2.05%**

Performance on Public Datasets

Dataset	Subset	Model	RMSE	MAE	WMAPE
Retail	All articles	DeepAR	204.68	51.53	0.43
		GraphDeepAR	**196.13**	**50.35**	**0.42**
	Cold starts	DeepAR			
		GraphDeepAR			
	Connected articles	DeepAR			
		GraphDeepAR			
	Top-100 articles	DeepAR			
		GraphDeepAR			
E-commerce	All articles	DeepAR	30.36	3.39	0.67
		GraphDeepAR	**20.65**	**3.08**	**0.61**
	Cold starts	DeepAR			
		GraphDeepAR			
	Connected articles	DeepAR			
		GraphDeepAR			
	Top-100 articles	DeepAR			
		GraphDeepAR			

Note: we define cold starts as articles with less than five demand lags at the time of the forecast. Connected articles are articles that have edges with other articles.

Mean RMSE uplift:

- **4% for retail**
- **32% for e-commerce**

Performance on Public Datasets

Dataset	Subset	Model	RMSE	MAE	WMAPE
Retail	All articles	DeepAR	204.68	51.53	0.43
		GraphDeepAR	**196.13**	**50.35**	**0.42**
	Cold starts	DeepAR	44.79	19.83	0.66
		GraphDeepAR	**41.78**	**18.84**	**0.63**
	Connected articles	DeepAR	207.12	52.34	0.42
		GraphDeepAR	**198.46**	**51.12**	**0.41**
	Top-100 articles	DeepAR	419.40	171.28	0.36
		GraphDeepAR	**401.27**	**164.10**	**0.35**
E-commerce	All articles	DeepAR	30.36	3.39	0.67
		GraphDeepAR	**20.65**	**3.08**	**0.61**
	Cold starts	DeepAR	**8.66**	2.62	0.79
		GraphDeepAR	8.72	**2.62**	**0.79**
	Connected articles	DeepAR	31.40	3.59	0.69
		GraphDeepAR	**21.40**	**3.18**	**0.61**
	Top-100 articles	DeepAR	164.68	42.78	0.98
		GraphDeepAR	**110.50**	**29.60**	**0.68**

Note: we define cold starts as articles with less than five demand lags at the time of the forecast. Connected articles are articles that have edges with other articles.

Mean RMSE uplift:

- **4% for retail**
- **32% for e-commerce**

Benefiting groups:

- **connected articles**
- **top-100 articles**

Running Time Difference

Dataset	Model	Training time	Inference time	Total difference
Retail	DeepAR	10.80 min	0.14 min	160.96%
	GraphDeepAR	28.33 min	0.22 min	
E-commerce	DeepAR	90.28 min	3.26 min	154.64%
	GraphDeepAR	234.73 min	3.46 min	
adidas	DeepAR	55.92 min	20.69 min	120.28%
	GraphDeepAR	139.80 min	28.96 min	

- Article similarity is calculated and stored before training
- Training is slower due to the need to backpropagate through graphs
- Inference speed of GraphDeepAR is comparable

Summary

- Incorporating article relationships in demand forecasting is challenging
- Our graph-based solution can address this challenge
    - Data-driven graph construction based on article attribute similarity
    - Integrates GNN encoder into the DeepAR forecasting model
    - Supports probabilistic forecasts
- Experimental results show that GraphDeepAR performs well
    - 2% financial uplift on adidas datasets
    - Up to 32% RMSE uplift on public datasets

Appendix

References

1. Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, *36*(4), 1420-1438.

2. Gandhi, A., Aakanksha, Kaveri, S., & Chaoji, V. (2021, September). Spatio-temporal multi-graph networks for demand forecasting in online marketplaces. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 187-203). Cham: Springer International Publishing.

3. Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211-243

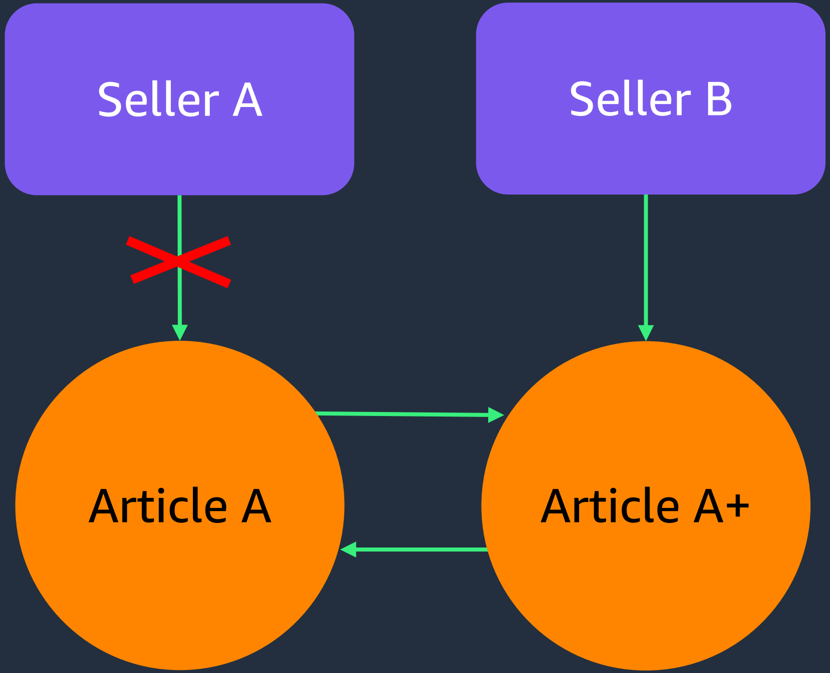
4. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181-1191.

5. Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.

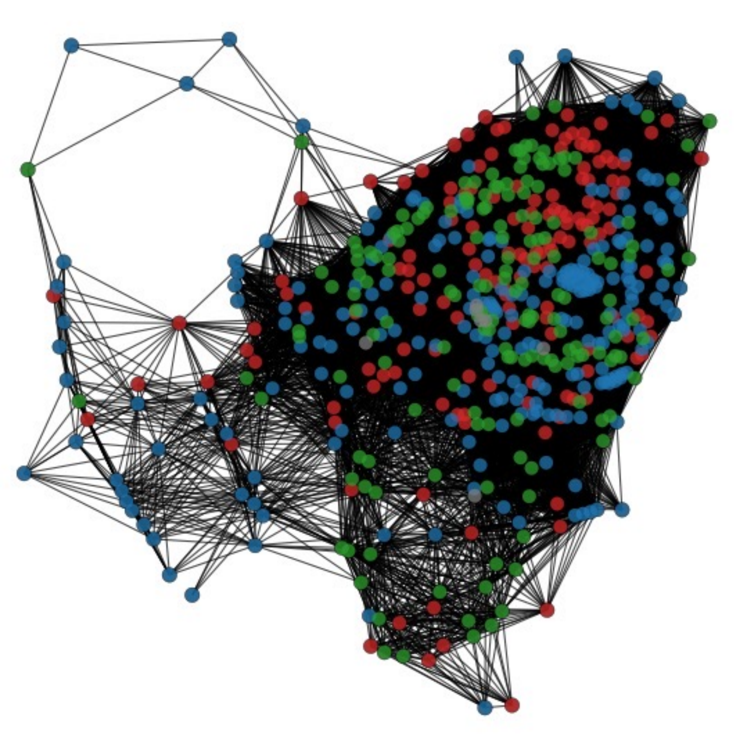
6. Veličković, P. (2023). Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, *79*, 102538.

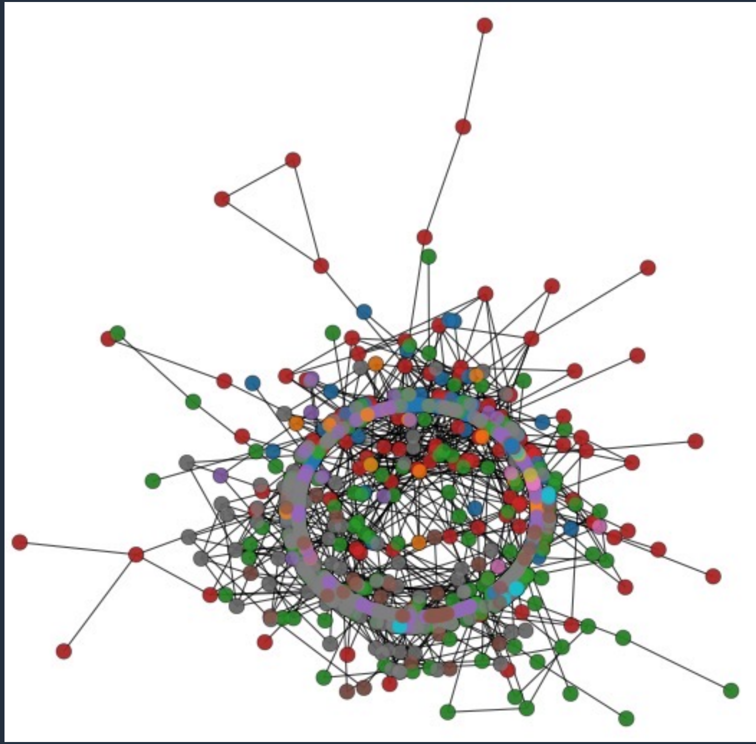
Importance of Article Relationships

- Out of stock status for the same article from other sellers
- Launch of competing articles
- Price change on a similar article by other sellers
- Sudden change in competitor‘s performance



Graph Illustration

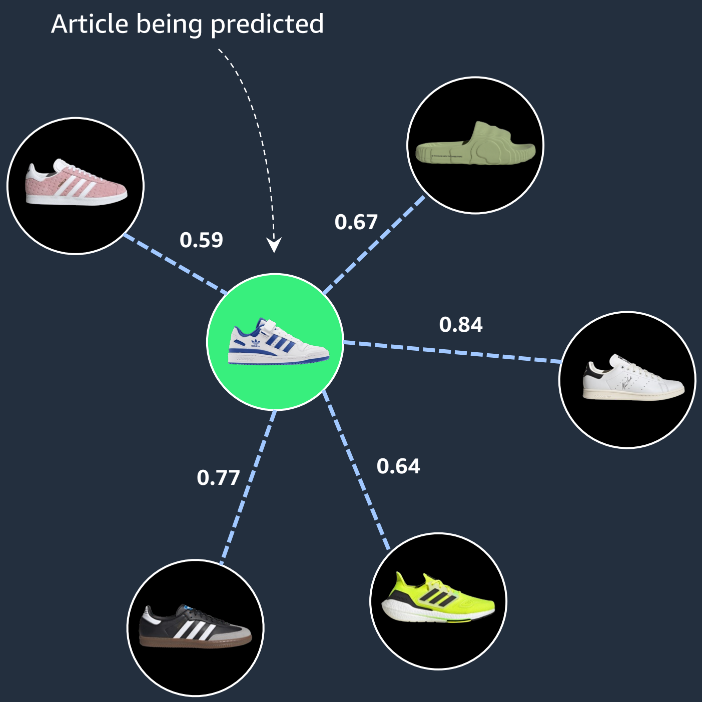




Article graphs for *Retail* dataset (left) and *E-commerce* dataset (right).

Sampling Mechanism [1/2]

- Graph contains thousands of articles
  - Average number of neighbors is high
  - Aggregating neighbors data is costly



Note: all numbers on this slide are artificial

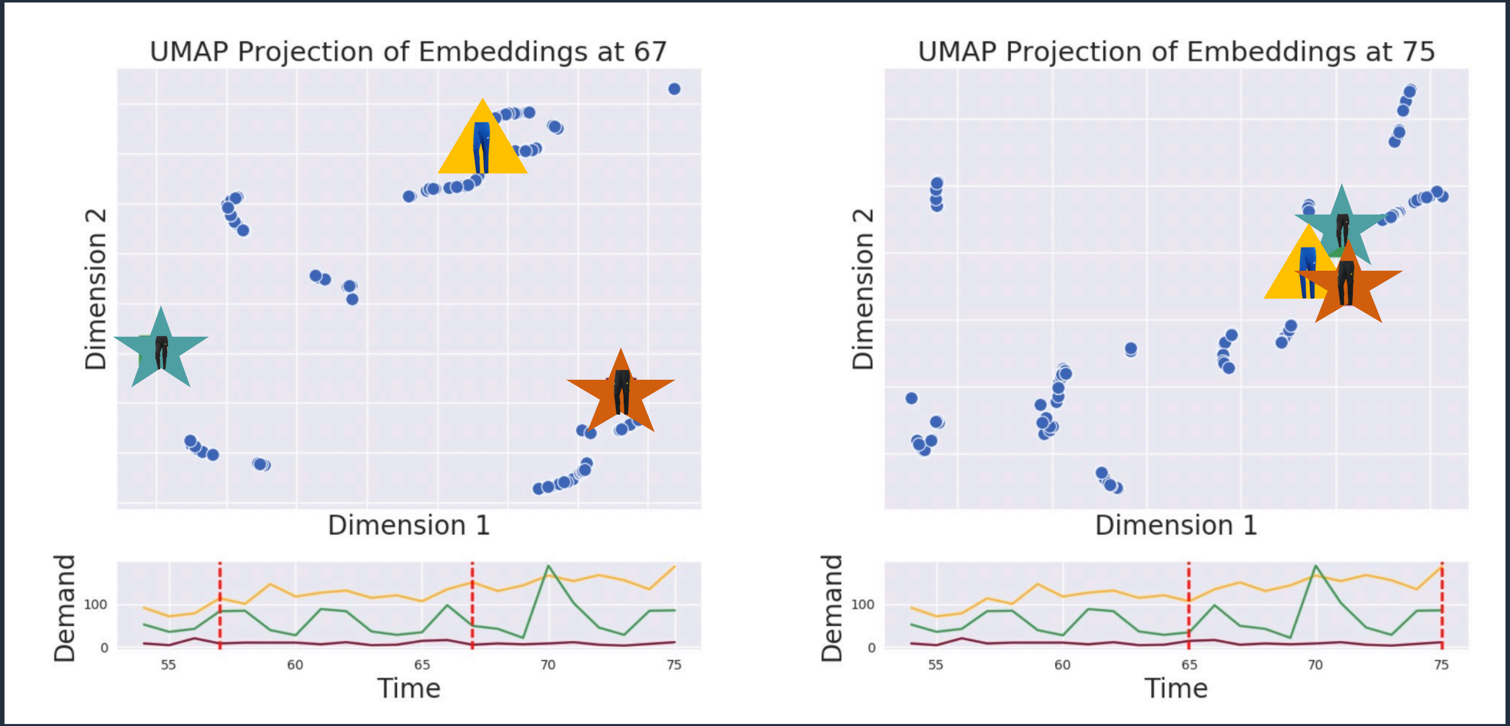
Sampling Mechanism [2/2]

- Graph contains thousands of articles
  - Average number of neighbors is high
  - Aggregating neighbors data is costly
- Solution: randomly sample neighbors
  - Different subset on each epoch
  - Helps scaling the solution

Note: all numbers on this slide are artificial

Time-Varying GNN Embeddings

Week 67 | Week 75



Meta-Parameters (Retail Dataset)

Dataset	Component	Meta-parameter	DeepAR	GraphDeepAR
retail	Sequential model	No. layers	2	2
		Hidden size	[128, 128]	[128, 128]
		Cell type	LSTM	LSTM
		Dropout	0.2	0.2
		Context length	10	10
	GNN encoder	No. layers	–	2
		Hidden size	–	[16, 8]
		Cell type	–	GCN
		Dropout	–	0.2
		Similarity cutoff	–	0.95
		Max no. neighbors	–	10
		Context length	–	10
	Training procedure	Max no. epochs	50	50
		Early stopping	5	5
		Learning rate	5×10^{-3}	5×10^{-3}
		Optimizer	Ranger	Ranger
		Loss function	t-distribution	t-distribution
		Batch sampler	Random	Synchronized

Example Predictions (Retail Dataset)