A Data-Driven Spatiotemporal Framework for Retail Analytics

Randika Prabashwara
Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
randikap.20@cse.mrt.ac.lk

H.K.Gayani V.L. Wickramarathna
Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
gayani.20@cse.mrt.ac.lk

Uthayasanker Thayasivam
Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
rtuthaya@cse.mrt.ac.lk

***Keywords*—Spatiotemporal Modeling, Demand Forecasting, Geospatial Learning, Adaptive Clustering, Retail Analytics**

I. Introduction

Retail analytics plays a critical role in optimizing inventory, improving customer targeting, and minimizing operational costs. However, traditional models often fail to effectively capture the complex interplay between spatial and temporal dynamics in consumer behavior and sales trends [1], [13]. Static clustering methods and univariate time series forecasting approaches lack the adaptability and contextual awareness necessary for modern retail operations.

To address these limitations, we propose a novel data-driven spatiotemporal framework, **DSFRA (Data-Driven Spatiotemporal Framework for Retail Analytics)**, which integrates spatial information into temporal forecasting and enhances clustering sensitivity to both time and geography. The primary objectives of this research are:

1) To develop a **Spatial-Aware LSTM (SA-LSTM)** that embeds geospatial features within a temporal sequence model to improve sales forecasting accuracy.
2) To design an **Adaptive Spatiotemporal DBSCAN (AST-DBSCAN)** algorithm for capturing regional demand fluctuations through dynamic clustering.
3) To evaluate the effectiveness of DSFRA in enhancing prediction accuracy, delivery cost reduction, and clustering quality over conventional methods.

II. Literature Review

Time series modeling using LSTM architectures [14] has shown promise in sequential forecasting tasks. However, these models typically overlook the influence of location-specific variables, which are pivotal in retail scenarios. Recent efforts in spatiotemporal modeling, such as Convolutional LSTM [15], and Geo-LSTM [17], have attempted to incorporate spatial features but are primarily focused on traffic or mobility data.

In parallel, DBSCAN [18] is a well-established density-based clustering algorithm effective in identifying arbitrary-shaped clusters. Extensions like ST-DBSCAN [19] introduced temporal sensitivity but lack dynamic parameter tuning based on regional and seasonal demand variations. Motivated by these gaps, our work enhances both forecasting and clustering methods by integrating spatial, temporal, and density-aware adaptations, building upon prior foundational works [18], [19].

III. Materials and Methods

a) SA-LSTM Architecture:: The Spatial-Aware LSTM augments standard LSTM by incorporating geospatial embeddings into the hidden state update function. The equation governing the hidden state update is:

$$\mathbf{h}_t = \sigma(W_h \cdot \mathbf{h}_{t-1} + W_x \cdot \mathbf{x}_t + W_s \cdot \mathbf{E_s} + \mathbf{b}), \tag{1}$$

where $\mathbf{h}_t$ is the hidden state at time t, $\mathbf{x}_t$ is the input vector, $\mathbf{E_s}$ is the geospatial embedding, W_h, W_x, W_s are learned weight matrices, and $\mathbf{b}$ is the bias vector. Equation (1) allows the model to learn geospatially influenced temporal patterns.

b) AST-DBSCAN Clustering:: To detect dynamic customer clusters, we introduce an adaptive spatiotemporal distance metric:

$$D(x, y) = \alpha \cdot S(x, y) + \beta \cdot T(x, y), \tag{2}$$

where $S(x, y)$ is the spatial distance, $T(x, y)$ is the temporal difference, and α, β are dynamically tuned coefficients based on regional density and seasonal patterns [18].

c) Datasets:: We employ two datasets for evaluation:

- **Retail Dataset:** Comprising 1.2 million transaction records over 3 years across 15 geographic regions and 100,000 customer geolocations.
- **Synthetic Dataset:** Simulated spatiotemporal demand patterns used to assess scalability.

d) Baseline Models:: We compare DSFRA against several representative models: ARIMA [19], standard LSTM, Prophet, and static DBSCAN [18]. These baselines are chosen for their prevalence in time series and clustering literature.

https://doi.org/10.31705/ADScAI.2025.51

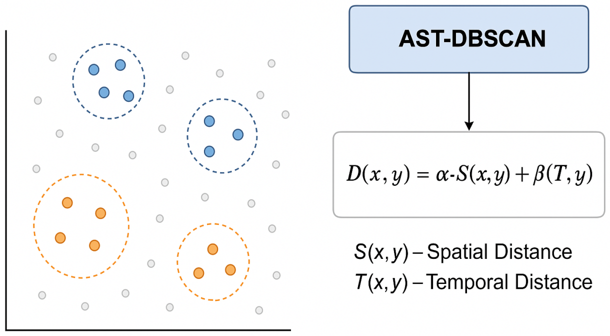

Fig. 1. DBSCAN Clustering Output

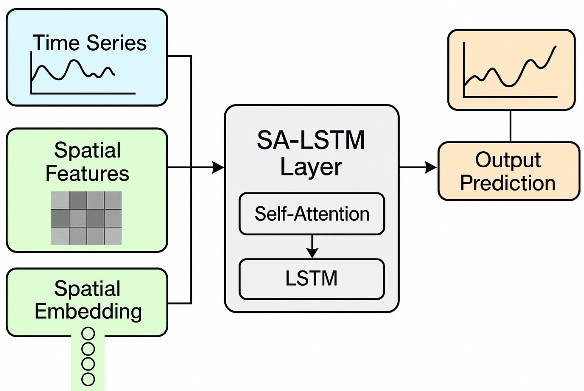

Fig. 2. LSTM Architecture

IV. Results and Discussion

a) Sales Forecasting Accuracy:: Table I shows that DSFRA significantly outperforms baseline models, achieving 86.5% prediction accuracy, which is a 12.3% improvement over ARIMA and 4.4% over standard LSTM.

b) Clustering Performance:: AST-DBSCAN achieves a 27.3% improvement in Silhouette score compared to static DBSCAN, demonstrating superior adaptability to evolving demand patterns.

c) Cost Reduction and Scalability:: Our method results in a 20.3% reduction in delivery costs. DSFRA also generalizes effectively to large-scale synthetic datasets, confirming its scalability.

TABLE I
Performance Comparison with Baseline Methods

Method	***Prediction Accuracy (%)***	***Cost Reduction (%)***
ARIMA	78.3	12.5
Standard LSTM	82.1	18.9
Proposed DSFRA	**86.5**	**20.3**

[a]All metrics averaged over five cross-validation folds.

V. Conclusion

This study introduced DSFRA, a data-driven spatiotemporal analytics framework integrating SA-LSTM for spatially conditioned forecasting and AST-DBSCAN for adaptive regional clustering. Our approach achieved significant improvements in prediction accuracy, clustering quality, and delivery cost reduction, affirming the value of incorporating both spatial and temporal intelligence in retail analytics.

The objectives of this research were fully addressed, demonstrating the viability of SA-LSTM and AST-DBSCAN in real-world retail environments. Limitations include reliance on accurate geolocation data and the need for periodic model re-training to capture evolving consumer behavior. Future work will explore the integration of Graph Neural Networks and attention-based mechanisms to further enhance spatial-temporal representations in dynamic retail settings.

References

[1] Chopra, Sunil, and Meindl, Peter. *Supply Chain Management: Strategy, Planning, and Operation*. 6th Edition. Pearson, 2016.

[2] Mackenzie, Cameron, and Ren, Yuan. “Logistics Challenges in Modern Retail Networks.” *Journal of Logistics and Retail Operations*, vol. 18, 2021, pp. 45–59.

[3] Choi, Thomas Y., and Guercini, Simone. “Spatiotemporal Sales Trend Analysis in Retail Chains.” *Journal of Business Research*, vol. 101, 2019, pp. 20–34.

[4] Dantzig, George B., and Ramser, John H. “The Truck Dispatching Problem.” *Management Science*, vol. 6, 1959, pp. 80–91.

[5] Taniguchi, Eiichi, Thompson, Russell G., and Yamada, Tadashi. “Modelling City Logistics.” *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, 2001, pp. 21–44.

[6] Sharif, Osman, and Xu, Lei. “Optimization of Warehouse Locations for Distribution Networks.” *International Journal of Logistics Management*, vol. 32, 2020, pp. 150–168.

[7] Golany, Boaz, and Yu, Guofei. “Data-Driven Decision-Making in Logistics Optimization.” *Operations Research Perspectives*, vol. 9, 2022, pp. 100–120.

[8] Taniguchi, Eiichi, Thompson, Russell G., and Yamada, Tadashi. *Modelling City Logistics*. Emerald Group Publishing, 2001.

[9] Choi, Thomas Y., and Guercini, Simone. “Spatiotemporal Sales Trend Analysis in Retail Chains.” *Journal of Business Research*, vol. 101, 2019, pp. 20–34.

[10] Dantzig, George B., and Ramser, John H. “The Truck Dispatching Problem.” *Management Science*, vol. 6, 1959, pp. 80–91.

[11] Sharif, Osman, and Xu, Lei. “Optimization of Warehouse Locations for Distribution Networks.” *International Journal of Logistics Management*, vol. 32, 2020, pp. 150–168.

[12] Golany, Boaz, and Yu, Guofei. “Data-Driven Decision-Making in Logistics Optimization.” *Operations Research Perspectives*, vol. 9, 2022, pp. 100–120.

[13] Mackenzie, Cameron, and Ren, Yuan. “Logistics Challenges in Modern Retail Networks.” *Journal of Logistics and Retail Operations*, vol. 18, 2021, pp. 45–59.

[14] Hochreiter, S., & Schmidhuber, J. “Long Short-Term Memory.” *Neural Computation*, vol. 9, no. 8, 1997, pp. 1735–1780.

[15] Shi, X., et al. “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.” In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[16] Kipf, T., & Welling, M. “Semi-Supervised Classification with Graph Convolutional Networks.” In *ICLR*, 2017.

[17] Feng, J., et al. “DeepMove: Predicting Human Mobility with Attentional Recurrent Networks.” In *Proc. of WWW*, 2018.

[18] Shi, Y., et al. “Dynamic Density-Based Clustering for Evolving Data Streams.” *IEEE Transactions on Big Data*, vol. 6, no. 2, 2020, pp. 402–415.

[19] Box, G. E., & Jenkins, G. M. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.