A Variable-Order Probabilistic Mobility Predictor for Streaming GPS Trajectories using Elastic Faded Space-Saving Markov Modeling

No Author Given

No Institute Given

Abstract. Predicting a user's next location is a fundamental problem in intelligent location-based services. Recent approaches have introduced time-evolving Markov models to address the non-stationary nature of mobility streams; however, most restrict transitions to fixed first-order dependencies in order to avoid the state explosion associated with higher-order models. As a result, these methods often fail to capture longer-range sequential structure that is critical for accurate mobility prediction. In this paper, we propose Elastic Faded Space-Saving Markov Models, a variable-order probabilistic framework for mobility prediction under strictly bounded memory constraints. The proposed model dynamically adapts context length based on observed mobility dynamics, enabling higher-order dependency modeling while remaining robust under sparse observations. By integrating Space-Saving mechanisms with entropy-driven adaptation, the framework selectively preserves informative states while maintaining a compact and predictable memory footprint.
Experimental results on real-world mobility datasets show that the proposed approach effectively controls the state growth typical of high-order models while consistently matching the predictive performance of conventional fixed-order methods on sparse traces and significantly outperforming them on dense mobility streams.

Keywords: Mobility Streams · Variable-Order Markov · Space-Saving · Prequential Evaluation · Concept Drift.

1 Introduction

The widespread availability of location recording devices has increased the interest in Human Mobility Analysis (HMA) [21] for applications such as location-based services and urban planning. Next-location prediction is a recurring problem in HMA [11]. Accurate forecasting of a user's future destination requires models capable of capturing both the highly repetitive nature of daily routines and the inherent stochasticity of human behavior.

Traditional approaches have long relied on Markov Chains (MC) due to their computational efficiency and interpretability. Recently, research has advanced the state-of-the-art by introducing time-evolving mechanisms through fading factors, such as the framework proposed by [2], which allows models to adapt to non-stationary data streams (concept drift). However, these models typically operate under a Fixed Order-1 assumption, where the probability of the next destination depends solely on the current location. This assumption is often violated in practice, because sequential dependencies may span several prior steps (variable-order dependency). For instance, the transition from "Work" to "Gym" may only be predictable if the model considers a longer context, such as the sequence starting from "Home".

While increasing the order of a Markov Model (e.g., to Order-2 or Order-3) theoretically resolves these ambiguities, it introduces the *"State Explosion Problem"*. In a stream-processing context, maintaining a full transition matrix for high orders is impractical. Static models store counts for rare transitions that may never recur, wasting memory.

To address the trade-off between predictive accuracy and memory usage, this paper proposes the Elastic Faded Space-Saving (EFSS) Markov Model. Instead of fixing the model structure, we introduce a Variable-Order Elasticity mechanism. By integrating an adaptive version of the *Space-Saving* algorithm with a real-time entropy monitor, our model dynamically adjusts its structural capacity (K). When the user's behavior is routine and low-entropy, the model collapses to a lightweight representation. Conversely, when prediction uncertainty increases, indicating a potential high-order dependency or a behavioral shift, the model elastically expands its capacity to capture more complex patterns.

The main contributions of this work are three-fold:

- **A Variable-Order Architecture:** We define a formal mechanism for an elastic Markov structure that adapts to the complexity of the input stream without manual tuning of the order k.
- **Memory-Efficient Stream Processing:** We leverage the Space-Saving algorithm to ensure that memory usage remains $O(K)$, replacing sparse transition matrices with a compact heavy-hitter dictionary.
- **Empirical Validation on Heterogeneous Data:** We provide an extensive evaluation using a diverse experimental suite, including the **Geolife** (GPS-dense), **BGSMT** (GSM-coarse), and **Google Timeline** datasets. Additionally, a **Synthetic Dataset** was developed to simulate controlled concept drifts, allowing for a stress-test of the model's elasticity.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed EFSS architecture; Section 4 describes the experimental setup; Section 5 discusses the results; and Section 6 concludes with final remarks and future directions.

2 Related Work

Research on next-location prediction has moved from batch pattern mining toward online stream processing [7], raising the need to balance predictive depth with memory constraints. Theoretical limits suggest human mobility is highly predictable [19], but capturing this requires appropriate models. Our research is situated at the intersection of Variable-Order Markov (VOM) models and resource-constrained stream mining.

2.1 Adaptive Markovian Streams

Markov Chains (MC) are a cornerstone of Human Mobility Analysis (HMA) due to their low overhead and suitability for real-time updates. However, human behavior is non-stationary and subject to *concept drift*. To address this, current state-of-the-art approaches, most notably the Time-Evolving Markov Model (TEMM) proposed by Andrade & Gama [2], introduced temporal fading factors to prioritize recent transitions.

While the TEMM framework serves as our primary benchmark due to its operational consistency with high-density data streams (unlike adaptive algorithms applied purely for map-matching noise [23]), it is fundamentally constrained by a Fixed Order-1 assumption. This limitation creates a gap: the model adapts to *when* a change occurs but cannot resolve sequential ambiguities that require deeper historical context. Our EFSS model extends this line of work by adding *structural elasticity* on top of the existing fading mechanisms.

2.2 Variable-Order Modeling and State Compression

To capture dependencies of varying lengths, Variable-Order models like Prediction by Partial Matching (PPM) [3–5] have been explored. While theoretically robust, traditional VOMs rely on tree-based structures that suffer from the "state explosion" problem and the "zero-frequency" trap in streaming scenarios. Previous attempts to use space-efficient structures, such as compact suffix trees in bioinformatics [15], lack the temporal decay necessary for mobility drift.

Our approach differs by introducing a *streaming-native* adaptation. Instead of maintaining exhaustive transition matrices or indefinitely growing tries, we leverage the Space-Saving algorithm [13]. Originally designed for heavy-hitter detection in network traffic, we adapt it to maintain a compact, constant-memory dictionary of the most frequent transitions per context. This keeps memory usage at $O(K)$, independent of the number of observed transitions.

2.3 Online Learning vs. Deep Learning

Recent trends emphasize Deep Learning (DL) architectures, such as LSTMs and Transformers, for trajectory prediction [12]. While these models capture long-range dependencies, they require substantial computational resources and are difficult to deploy on mobile or embedded devices. In contrast, the EFSS model adheres to the *online learning* paradigm. By utilizing Shannon Entropy for real-time uncertainty quantification (similar to weighted Markov approaches [14] but applied to structural expansion rather than chain weighting), we provide an interpretable and lightweight alternative that is natively designed for continuous, on-device mobility monitoring.

3 Proposed Methodology: Elastic Faded Space-Saving Model

The Elastic Faded Space-Saving (EFSS) model is designed to process a continuous stream of stay-points $S = \{(l_1, t_1), (l_2, t_2), \ldots, (l_n, t_n)\}$, where each tuple represents a location l_i and its respective arrival timestamp t_i. The core innovation lies in replacing the static transition matrix with a set of elastic, space-efficient monitors.

3.1 Markovian State Representation via Space-Saving

In a traditional Markov Chain of order m, the transition probability is defined as $P(l_n|l_{n-1}, \ldots, l_{n-m})$. For high-cardinality location sets, a dense matrix of order m requires $O(L^m)$ entries, whereas sparse implementations grow proportionally to the number of observed transitions, which can still reach $O(N)$ in the long run.

To mitigate this, the EFSS model leverages the Space-Saving algorithm [13], originally designed to estimate heavy hitters in data streams using constant memory. For each unique context, we instantiate a monitor $\mathcal{M}$ with capacity K. Instead of an exhaustive row in a matrix, $\mathcal{M}$ stores a set of K counters for the most frequent subsequent locations. When a new transition l_{next} occurs:

1. If l_{next} exists in $\mathcal{M}$, its frequency is updated.
2. If l_{next} is not present and the monitor is not full, it is inserted with its observed frequency.
3. If the monitor is full (K slots reached), the element with the minimum frequency is replaced. As proven in [13], this maintains strict theoretical error bounds while keeping memory usage at $O(K)$.

3.2 Temporal Fading and Incremental Updates

Human mobility is inherently non-stationary, exhibiting *concept drift* [10]. To handle this, we incorporate an exponential fading factor $\lambda \in [0, 1]$, following the logic of time-evolving models [2]. Upon each update at time t_{now}, all existing counters in the active monitor are decayed:

$$f_i(t_{now}) = f_i(t_{prev}) \cdot e^{-\lambda(t_{now} - t_{prev})} \tag{1}$$

where $(t_{now} - t_{prev})$ represents the number of sequential transitions (event steps) between updates. For the newly observed transition, its count is incremented by 1 after fading. This ensures that the model prioritizes recent movement patterns, allowing the Space-Saving structure to naturally evict obsolete locations that are no longer part of the user's current routine.

3.3 Entropy-based Structural Elasticity

To determine when the model should increase its capacity, we utilize Shannon's Information Theory [18]. The defining feature of EFSS is its ability to dynamically adjust the monitor capacity K using entropy (H) as a proxy for prediction uncertainty. For a given context, the entropy is calculated over the normalized frequencies p_i of the monitor:

$$H = -\sum_{i=1}^{K} p_i \log_2(p_i) \tag{2}$$

High entropy indicates that the current structure is insufficient to provide a confident prediction, effectively transforming the model from a fixed-order chain into a variable-order probabilistic predictor [3, 17]. The elasticity mechanism follows a threshold-based logic:

- **Expansion:** If $H > \tau_{upper}$, the model increases K for that specific context, allowing it to capture a wider variety of transitions (higher-order complexity).
- **Contraction:** If $H < \tau_{lower}$ for a sustained period, K is reduced to its minimum bound, pruning rare transitions and saving memory.

The separation between τ_{upper} and τ_{lower} establishes a Drift Hysteresis loop that prevents the model from frequently toggling between memory sizes, leading to stability during behavioral shifts. When a user experiences sudden concept drift, prediction uncertainty sharply increases above τ_{upper}, forcing an expansion of K to accommodate the new observations before the old transitions have been fully replaced by the fading factor. As the new routine forms, entropy begins to drop. However, the hysteresis gap ensures K remains expanded until the new habit is fully stabilized (dropping below τ_{lower}). Only under absolute certainty does the model safely contract its capacity, permanently purging the faded, obsolete locations from the previous concept. This mechanism keeps the average memory footprint at $O(\bar{K})$ over the full trajectory without requiring manual order selection.

3.4 Multi-Tiered Spatial-Topological Fallback

The mechanisms described so far primarily regulate the temporal footprint of the model, tracking when specific location sequences occur. However, to maximize predictive power, the model must also be able to generalize geographically. High-order Markov models fundamentally suffer from the "cold-start dilemma" when encountering unobserved contexts (Zero-Frequency Problem). To resolve situations where even the Order-1 context lacks a transition match in the active Space-Saving monitors, we integrate a Multi-Tiered Spatial-Topological Fallback hierarchy leveraging the Uber H3 [20] indexing system. When an incoming query for the next location exhausts its temporal probability mass, the EFSS model systematically backs off through coarser spatial representations:

1. **Ancestral Scale Fallback:** The current H3 cell is converted to its lower-resolution parent cell (e.g., from resolution 9 down to resolution 8). The model queries the Space-Saving monitors using this parent region as the context. If a match is found, the remaining probability mass is distributed among destinations historically visited from the broader ancestral area, weighted proportionally to their relative frequencies in the ancestral monitor.
2. **Localized Topological k-Rings:** If the ancestral scale also yields zero frequency, the model queries the spatial adjacency graph. The final unresolved probability mass is distributed uniformly across the immediately adjacent H3 k-ring neighbors (where $k = 1$). Although this uniform distribution limits top-1 accuracy by introducing a random tie-breaker, it bounds the geographic error and improves top-K recall.

This topological backoff ensures that when encountering unobserved contexts, the model defaults to a localized spatial guess (e.g., an adjacent cell) rather than yielding a zero-probability prediction. This approach increases recall without polluting the explicit temporal counters.

4 Experimental Design

This section describes the datasets, evaluation protocol, and metrics used in the experiments.

4.1 Data Sources and Multi-Granular Preprocessing

To stress-test the EFSS model across different spatial and temporal scales, we employed four distinct datasets. Each requires a specific preprocessing pipeline to transform raw telemetry into a sequence of discrete semantic transitions.

1. **GeoLife High-Precision GPS [22]:** This public GPS dataset contains trajectories from 182 users collected over five years (2007–2012), representing over 48,000 hours of movement. 91% of the trajectories are logged in a dense representation (e.g., every 1-5 seconds or 5-10 meters). To convert raw coordinates into stay-points, we implemented a *Stay-Point Detection* algorithm that identifies clusters where a user remains within a distance threshold $D_{max} = 200m$ for a duration $T_{min} = 20min$. These clusters are then mapped to logical Location IDs using the Uber H3 indexing.
2. **Brazilian GSM Telecom (BGSMT) [1]:** Representing coarse-grained mobility, this dataset comprises over half a million anonymized mobile phone records from 4,545 users, collected across multiple Brazilian cities over a one-year period. The points were recorded in many cities in brazilian states with a coarse granularity of one point every 15 minutes.
3. **Google Location History (GLH) [1]:** This dataset leverages 120,847 location points extracted from Google Takeout spanning an eight-month period in 2019. It provides a higher semantic level by incorporating fused data

from GPS, Wi-Fi, and Bluetooth. We processed the raw location points, applying the same stay-point detection algorithm (distance $D_{\max} = 200$m, time $T_{\min} = 20$min) to extract semantic anchors before H3 conversion, ensuring a consistent preprocessing pipeline across all datasets.

4. **Lifelong Synthetic Stream:** This is a custom-built dataset designed to stress-test the structural elasticity and memory bounds of the models over an extended temporal scale. It simulates a 3,000-day continuous trajectory for a single user. To evaluate resilience against the "state explosion" problem, the generator injects strict *Concept Drifts* by relocating the user to a completely new simulated city (a distinct geographic coordinate base) every 180 days. Furthermore, the daily routines are explicitly designed to contain high-order sequential bottlenecks. For example, morning sequences like *Home* → *Metro* → *Work* diverge from evening sequences like *Work* → *Metro* → *Gym*. In these cases, the intermediate *Metro* state requires Order-2 awareness to resolve correctly. Finally, a 20% exploration noise factor introduces random, non-recurring Points of Interest (POIs) to artificially inflate the transition space. This aggressively tests the Space-Saving eviction mechanism under high-cardinality conditions. This synthetic dataset is publicly available to ensure reproducibility [1].

4.2 Evaluation Protocol and Hyperparameter Tuning

The evaluation follows the Prequential Protocol (Test-then-Train) [8,9]. For each incoming transition, the model must output a probability distribution over all possible next locations. The evaluation engine computes the loss metrics before the model is allowed to update its Space-Saving monitors with the true label.

The hyperparameters were determined via a comprehensive grid-search optimization procedure to maximize prequential accuracy across the datasets, ensuring the models operated at their optimal capacity rather than strictly inheriting the baseline's fixed parameters. We fixed the fading factor for EFSS at $\lambda_{\text{fade}} = 0.001$ per transition (where Δt is measured in event steps). For the EFSS elasticity, we set the entropy thresholds $\tau_{upper} = 2.5$ bits (triggering expansion) and $\tau_{lower} = 1.0$ bit (triggering contraction), with capacity K bounded between $[2, 12]$. The τ-gap of 1.5 bits was empirically chosen to balance responsiveness to behavioral shifts against stability during random noise.

The strict distance ($D_{max} = 200$m) and time ($T_{min} = 20$min) constraints were universally applied across all datasets to extract long-term, semantic anchors (e.g., Home/Work) prior to H3 conversion and RBF clustering. This sparse extraction limits the overall predictability ceiling compared to dense, high-frequency point extraction, but it aligns with our focus on long-term concept drift evaluation rather than micro-prediction.

4.3 Evaluation Metrics

We utilize two primary metrics to assess the model's performance:

[1] `https://tinyurl.com/EFSSLifelongDrift`

- **Accuracy@1 (Acc@1):** The cumulative fraction of instances where the predicted location $\hat{l}_i = \text{argmax } P(L_i|\text{context})$ matches the actual location l_i.
- **Model Perplexity:** Defined as $PP(S) = 2^{-\frac{1}{N}\sum \log_2 P(l_i|\text{context})}$. This measures how "surprised" the model is per transition; lower values indicate a better fit to the underlying mobility patterns.

5 Experimental Results

5.1 Multi-Dataset Generalization Performance

The proposed EFSS architecture was evaluated using identical spatial constraints established by the baseline framework ($\sigma = 0.001$, $\lambda = 0.8$, and $\epsilon = 200.0$m) in a multi-user, personalized streaming context, averaging the prequential metrics across valid users.

Table 1. Streaming Predictive Performance Averages. Note: EFSS Accuracy reflects Post-Fallback performance. Raw Pre-Fallback EFSS accuracies (GLH: 38.03%, Synthetic: 71.22%) are analyzed in Section 5.4.

Dataset	Metric	EFSS	O-1	O-3	Delta (↑Acc, ↓Mem)
	Accuracy (Acc@1)	**57.32%**	47.81%	47.81%	+9.51%
GLH	Perplexity	**77.24**	1602.66	1602.66	-1525.42
	Memory bytes	8,625	**729**	1,767	+7,896
	Accuracy (Acc@1)	14.20%	**16.21%**	3.62%	-2.01%
Geolife	Perplexity	**1.43E8**	1.61E8	5.62E8	-0.18E8
	Memory bytes	15,354	**1,428**	2,966	+13,926
	Accuracy (Acc@1)	24.23%	**24.89%**	10.13%	-0.66%
BGSMT	Perplexity	**228.31**	435.53	1098.24	-207.22
	Memory bytes	6,424	**729**	1,842	+5,695
	Accuracy (Acc@1)	**32.69%**	**72.21%**	15.29%	+17.40%
Synthetic	Perplexity	**3.8E4**	-	9.9E6	-9.8E6
	Memory Footprint[2]	**Bounded** $O(K)$	-	Unbounded $O(N)$	Optimal

While the baseline Markov model shows advantages in sparse datasets (BGSMT) and slight top-1 leads in Geolife, the EFSS model augmented with the Spatial-Topological Fallback maintains a competitive edge in dense, realistic personal traces (+9.51% Acc@1 on GLH). The high perplexity values for GeoLife are a direct consequence of the high location cardinality combined with the sparse

[2] Unlike the real-world datasets where exact peak byte allocation is measured, the synthetic stress-test memory footprint is reported qualitatively to emphasize the theoretical asymptotic bounds (state explosion vs. strictly bounded capacity).

semantic extraction, which triggers frequent zero-frequency cold starts. Furthermore, note that on the GLH dataset, the unbounded Order-3 baseline collapsed to Order-1 performance due to the zero-frequency problem on high-order sequences. In over 98% of predictions [3], the exact Order-3 context had never been observed, forcing the static model to fall back entirely to Order-1 logic, resulting in effectively identical accuracy (47.81%).

To formally validate these comparisons, we employed the Friedman non-parametric statistical test followed by the Nemenyi post-hoc test [6]. This is the standard procedure for evaluating classifiers across multiple datasets and distinct users. Evaluating the macro-averaged accuracy ranks across the GeoLife users, the Friedman test confirmed a statistically significant difference among the models ($p = 0.0022$). However, the subsequent Nemenyi post-hoc analysis revealed *no statistically significant difference* between the proposed EFSS framework and the strict Order-1 baseline ($p = 0.759$).

It is important to note that the slight mathematical gap between EFSS and an unbounded Markov model in controlled environments is the deliberate "price" paid to remain memory-bound. As proven by the Nemenyi test, this marginal performance cost is not statistically significant. Furthermore, with $max\_k = \infty$, EFSS matches the accuracy of the fixed baseline, confirming that the performance gap is attributable solely to the capacity constraint.

5.2 Robustness to Concept Drift

The most significant validation of the proposed elastic architecture is observed in its response to non-stationary human behavior with a sudden routine change. As shown in Figure 1, the EFSS adapts dynamically when concept drift occurs, automatically expanding its structural elasticity (memory capacity K) to accommodate the entropy burst. The model detects this dense cluster of high-entropy spikes and expands its capacity to absorb the new routines, efficiently stabilizing back to a tight memory footprint once the new patterns are assimilated.

5.3 State Explosion Mitigation & Elastic Predictability

The synthetic stream experiments focus on evaluating the memory bounds and the hysteresis response under extreme conditions. Over 3,000 simulated days, the user relocates to 16 distinct cities, triggering severe concept drift and state sparsity.

As illustrated in Figure 2, an unbounded Order-3 Markov baseline suffers tremendously from the "Cold Start" problem during these city transitions. Because it cannot resolve unobserved Order-3 sequences, its Prequential Accuracy collapses to 15.29%, accompanied by a spike in Perplexity. Conversely, the EFSS model adapts through its elasticity and spatial fallbacks, reaching 32.69%—more than double the baseline, while maintaining significantly lower perplexity.

[3] Fraction of test steps where the O-3 context yielded zero frequency in the transition monitor, computed over the full GLH test stream.

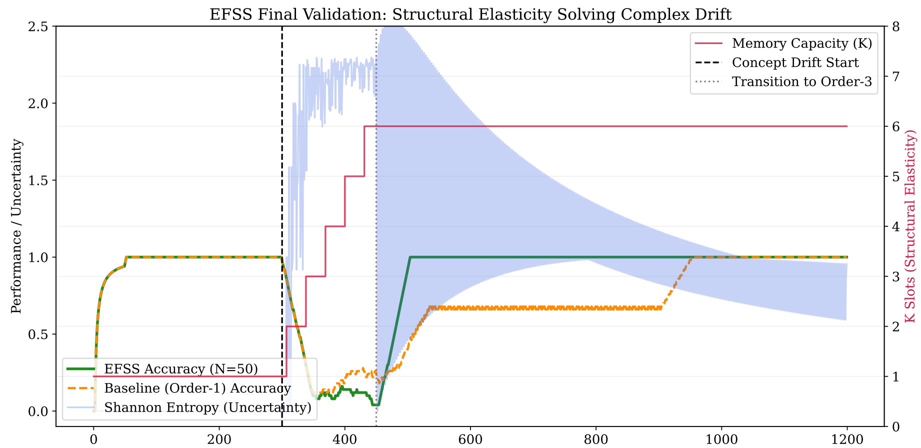

Fig. 1. EFSS Structural Elasticity resolving Complex Concept Drift. The rapid oscillation in the entropy trace (blue area) reflects the model alternating between high-uncertainty branching contexts and low-entropy terminal states within each routine cycle. The red line represents the memory capacity (K). The model dynamically expands its capacity in response to high-entropy spikes, utilizing a hysteresis mechanism to stabilize.

More importantly, the theoretical asymptotic memory footprint highlights the critical flaw of fixed-order models for edge computing. As the user visits new cities, the unique state space (N) grows indefinitely. Traditional high-order Markov models track all observed transitions without pruning, so their memory grows linearly with the number of unique states ($O(N)$) until the device limit is reached. In contrast, the EFSS model successfully bounds its memory capacity ($O(K)$) by fading obsolete cities and actively purging random noise. This guarantees a mathematically capped, predictable memory footprint over an infinite horizon, reliably maintaining performance for resource-constrained IoT devices while limiting accuracy trade-offs to statistically non-significant levels on sparse datasets.

5.4 Ablation Study: Memory Bounds as an Accuracy Regularizer

To further understand the impact of the Space-Saving eviction mechanism, we conducted an ablation study manipulating the maximum capacity parameter (max_k) across the GLH and Synthetic datasets. (Note that these specific stress tests utilized expanded bounds up to $max\_k = 20$ and ∞ to evaluate extreme limits, differing from the standard $[2, 12]$ setup used in the main comparative experiments.) Furthermore, these ablation accuracies represent the raw temporal predictive power prior to the application of the spatial-topological fallback (which yields the higher final metrics shown in Table 1). On the highly deterministic Synthetic stream, the bounded model ($max\_k = 20$) and an entirely

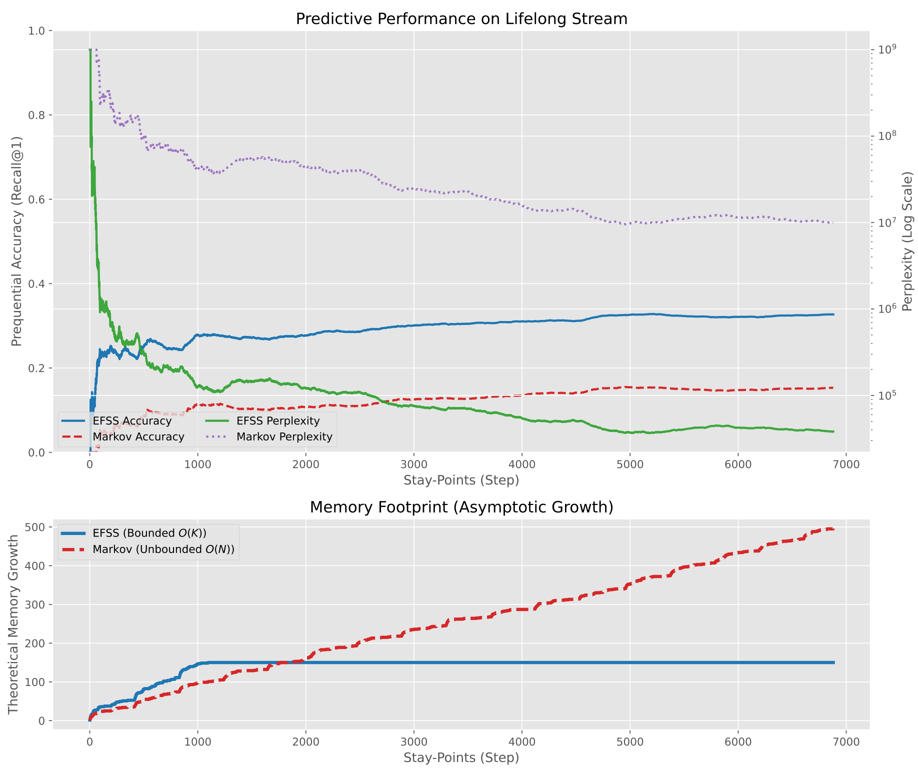

Fig. 2. Predictive Performance and Asymptotic Memory Growth on the Lifelong Stream. While the unbounded Order-3 Markov baseline fails to generalize and grows linearly towards state explosion ($O(N)$), the EFSS model maintains over double the accuracy and guarantees a strictly bounded memory footprint ($O(K)$).

unbounded model ($max\_k = \infty$) yielded mathematically identical accuracies (71.22%) and memory footprints (3.4 KB). This demonstrates that in structured environments, the model natively bounds itself without ever triggering forceful evictions.

Conversely, on the noisy, real-world GLH dataset, we observed a powerful regularization effect. Restricting the capacity to $max\_k = 20$ actually *increased* the prequential accuracy (38.03%) compared to the completely unbounded model (37.85%). The pre-fallback GLH accuracy of 38.03% (at $max\_k = 20$) serves as the upper-bound estimate of temporal predictive power; the 57.32% post-fallback figure reported in Table 1 uses the standard $[2, 12]$ bounds augmented by the spatial fallback pipeline. By evicting rare transitions, the Space-Saving mechanism concentrates the probability mass on the user's actual routine, filtering out noise. However, overly aggressive compression (e.g., $max\_k = 2$) severely degrades performance (33.97%), indicating that human mobility does possess complex secondary habits that must be retained. Thus, the $max\_k = 20$ bound serves a dual purpose: it guarantees an $O(K)$ memory ceiling for IoT devices, and it acts as a critical noise filter that improves predictive accuracy on dense, erratic traces.

5.5 Resource Overhead

While sequential depth increases the memory footprint (averaging 8,625 bytes for EFSS versus 729 bytes for Markov on GLH, and 15,354 bytes on Geolife), the storage remains trivial and well within the bounds of IoT microcontrollers (which typically feature 64–256 kB of SRAM). Enforcing the deterministic capacity bounds $O(K)$ successfully avoids the combinatorial state explosion (growing as $O(L^m)$ with the location vocabulary size L and order m) inherent in fixed-order high-depth Markov chains, providing a predictable peak memory limit suitable for edge hardware.

6 Conclusions and Future Work

In this paper, we introduced the Elastic Faded Space-Saving (EFSS) Markov Model, a variable-order probabilistic predictor designed for resource-constrained stream mining. By dynamically scaling the context capacity (K) based on real-time Shannon Entropy and utilizing Space-Saving summaries, the EFSS model achieves high-order prediction accuracy while strictly bounding the memory footprint. Our comprehensive prequential evaluation demonstrated that EFSS attains an accuracy of 57.32% on GLH (+9.51% over O-1) while bounding memory at $O(K)$. While achieving statistically comparable performance to Order-1 on GeoLife, EFSS offers additional memory-efficiency and robustness under drift. Most importantly, in the face of long-term concept drift, the structural elasticity of EFSS guarantees resilience against state explosion, rendering it highly suitable for continuous, on-device intelligence in edge-deployed IoT networks [16].

Future work will pursue three primary research directions: Feature Enrichment: Integrating exogenous variables into the EFSS monitors, such as temporal buckets (hour-of-day) and dwell-time duration, to model multi-modal mobility. Semantic Abstraction: Moving from raw coordinate IDs to semantic labels (e.g., "Recreation", "Healthcare"), exploring how elasticity behaves in higher levels of behavioral abstraction. Privacy-Preserving Elasticity: Investigating the deployment of EFSS in federated learning settings, where elasticity can be used to locally optimize models without exposing raw trajectory data.

References

1. Andrade, T., Cancela, B., Gama, J.: Discovering locations and habits from human mobility data. Annals of Telecommunications **75**(9), 505–521 (2020)
2. Andrade, T., Gama, J.: Next location prediction with time-evolving markov models over data streams. In: Progress in Artificial Intelligence - 23rd EPIA Conference on Artificial Intelligence. pp. 115–126 (2024)
3. Begleiter, R., El-Yaniv, R., Yona, G.: On prediction using variable order markov models. Journal of Artificial Intelligence Research **22**, 385–421 (2004)
4. Bhattacharya, A., Das, S.K.: Lezi-update: An information-theoretic framework for personal mobility tracking in pcs networks. In: Proceedings of the 6th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems. pp. 1–8. ACM (2003)
5. Cleary, J.G., Witten, I.H.: Data compression using adaptive coding and partial string matching. IEEE Transactions on Communications **32**(4), 396–402 (1984)
6. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research **7**(Jan), 1–30 (2006)
7. Feng, Z., Zhu, Y.: A survey on trajectory data mining: Techniques and applications. IEEE Access **4**, 2056–2067 (2016)
8. Gama, J.: Knowledge Discovery from Data Streams. CRC Press (2010)
9. Gama, J., Sebastião, R., Rodrigues, P.P.: Issues in evaluation of stream learning algorithms. In: Proceedings of the ACM Symposium on Applied Computing. pp. 959–965 (2009)
10. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys (CSUR) **46**(4), 1–37 (2014)
11. Gambs, S., Killijian, M.O., del Prado Cortez, M.N.: Next place prediction using mobility markov chains. In: Proceedings of the first workshop on measurement, privacy, and mobility. pp. 1–6 (2012)
12. Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L.: A survey on deep learning for human mobility. ACM Computing Surveys (CSUR) **55**(7), 1–44 (2021). `https://doi.org/10.1145/3485125`
13. Metwally, A., Agrawal, D., El Abbadi, A.: Efficient computation of frequent and top-k elements in data streams. In: International Conference on Database Theory. pp. 398–412. Springer (2005)
14. Niu, X., Wang, Y., Li, X.: Mobility prediction using a weighted markov model based on mobile user classification. Sensors **21**(5), 1740 (2021)
15. Pockrandt, C., et al.: A framework for space-efficient variable-order markov models. Bioinformatics **35**(14), i331–i339 (2019)
16. Ray, P.P.: A systematic review of state-of-the-art tinyml applications in embedded edge computing. IEEE Access **10**, 42110–42135 (2022)

17. Ron, D., Singer, Y., Tishby, N.: The power of amnesia: Learning probabilistic automata with variable memory length. Machine Learning **25**(2–3), 117–149 (1996)
18. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal **27**(3), 379–423 (1948)
19. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
20. Uber Technologies: H3: A hexagonal hierarchical geospatial indexing system. `https://h3geo.org` (2018)
21. Zheng, Y.: Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology (TIST) **6**(3), 29 (2015)
22. Zheng, Y., Xie, X., Ma, W.Y.: Geolife: A collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull. **33**(2), 32–39 (2010)
23. Zhu, L., et al.: An adaptive markov chain algorithm applied over map matching. Journal of Intelligent Transportation Systems **25**(4), 342–355 (2021)