# Topic 3: Deep Embedded Clustering using Multi-Modal Feature Fusion

January 4, 2026

## 1 Problem Formulation

Traditional snapshot models capture current profile state but miss the **trajectory**. The core challenge lies in distinguishing identical cross-sectional profiles with opposing vectors:

- **Ambiguity:** How to distinguish a loyal customer on a seasonal break from one about to churn?
- **Vector Divergence:** How to separate customers with similar profiles but different momentums?

We hypothesize that customer transitions do not lie on a linear plane, but on a discontinuous state space with specific transition probabilities defined by a "Customer DNA" embedding.

## 2 Methodology

### 2.1 Data Preparation

- **Static Features ($\mathcal{S}$):** Demographic and account metadata including Tenure, Income, Share of Wallet, Education, Location, and Loyalty Status.
- **Temporal Features ($\mathcal{T}$):** Sequential behavioral metrics tracking Frequency, Haul Length, Yield, Companion Ratio, and Points Accumulation, alongside seasonal indicators (Summer/Winter).
- **Velocity Vectors:** To support the velocity loss constraint, first-order derivatives were pre-computed for all temporal features (e.g., $\Delta\text{Freq} = \text{Freq}_t - \text{Freq}_{t-1}$).

### 2.2 Domain Spaces Embedding

We define the input space $\mathcal{X}$ as the product of static and temporal subspaces $\mathcal{X} = \mathcal{S} \times \mathcal{T}$, where $\mathcal{S} \in \mathbb{R}^{d_s}$ represents slowly changing demographics, and $\mathcal{T} \in \mathbb{R}^{d_t \times L}$ represents dynamic preferences. **The objective is to learn a non-linear mapping $\phi : \mathcal{X} \to \mathcal{Z}$ to a latent manifold using autoencoders.**

### 2.3 Hyperspherical Constraint

Euclidean distance fails in high-dimensional behavioral data. A corporate executive (High Spend/High Freq) and a consistent mid-tier business traveler may have vast Euclidean distances due to scale, but identical behavioral signatures.

We normalize embeddings to the unit hypersphere $\mathcal{S}^{d-1}$. Similarity is measured via cosine similarity:

$$\text{sim}(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{z}_1 \cdot \mathbf{z}_2 \quad \text{s.t.} \quad \|\mathbf{z}\| = 1 \tag{1}$$

This ensures that magnitude (total spend) does not overshadow the behavioral pattern (booking consistency).

### 2.4 The Physics-Informed Loss Function

To capture both state and trajectory, the loss function $\mathcal{L}$ incorporates four components:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{static}} + \beta\mathcal{L}_{\text{temp}} + \gamma\mathcal{L}_{\text{contr}} + \delta\mathcal{L}_{\text{vel}} \tag{2}$$

1. $\mathcal{L}_{\text{static/temp}}$: Reconstruction of **profile** and **forecast**.
2. $\mathcal{L}_{\text{contr}}$: Topological constraint ensuring semantic locality.
3. $\mathcal{L}_{\text{vel}}$: The differential penalty ($\frac{dx}{dt}$), learning the **velocity vector**—where the customer is going and how fast.

## 3 Latent Phase Space Clustering

Traditional clustering ($C_{\text{trad}}$) groups by position $\mu$. We introduce **Phase Space Clustering**, grouping by position *and* tangent vectors (velocity):

$$C_{\text{phase}} = \{(z, v) \mid d_{\mathcal{M}}(z, \mu_z) + \lambda\|v - \mu_v\| < \epsilon\} \tag{3}$$

This topologically separates a "Churner" from an "Acquired" user. Despite occupying the same spending coordinate $z$, their velocity vectors $v$ have a cosine similarity of $-1$.

## 4 Results: Cluster Profiles

Elbow method on Cosine Distance suggested $k = 4$. The clusters reveal distinct behavioral velocities.

Table 1: Cluster Analysis: State vs. Velocity

| Clust | Profile | State ($\sigma$) | Vel ($\Delta$) | Interpretation |
|---|---|---|---|---|
| **0** | Low Value<br>Decelerating | Freq: -0.34<br>Points: -0.48 | -0.30<br>Neg | **Active Churn**<br>Dying Momentum |
| **1** | Mass Mkt<br>Neutral | Freq: -0.21<br>Haul: -0.30 | +0.06<br>0.00 | **Static Orbit**<br>1-2x/year stable |
| **2** | High Value<br>Accelerating | Freq: +0.58<br>Yield: +0.66 | +0.31<br>+0.35 | **Expansion**<br>Business Growth |
| **3** | Vacation<br>Crash | Haul: +0.47<br>Comp: +0.69 | *N/A*<br>-0.21 | **Hidden Risk**<br>Post-Trip Crash |

**Cluster 0: Signature: Low state, negative velocity. Insight:** Momentum is too negative for gentle nudges.
**Action:** Ignore or perform shock offers.

**Cluster 1: Signature: Below avg state, zero velocity. Insight:** Behavior is rigid. Frequency uplift is unlikely.
**Action:** *Yield Maximization.* Focus on ancillary revenue (bags, seats) rather than frequency.

**Cluster 2: Signature: High state, positive velocity. Insight:** Naturally expanding wallet share.
**Action:** *Zero-Friction.* Focus on smooth customer experience, and priority handling to sustain the natural trajectory.

**Cluster 3: Signature: High yield spike, sharp negative velocity. Insight:** Often misclassified as Churning Business Travelers. High spend was event-driven (family trip).
**Action:** *Dream Trigger.* Ignore short-term frequency. Target next year's vacation window with "Dream Destination" offers.

## 5 Conclusion

1. **Autoencoder Dimensionality Reduction:** the model compressed complex sequential behaviors, velocity vectors and profiles into a letent 24 dimension space ($Z$) on a hyperspherical embedding

2. **Latent Space Clustering:** The application of Phase Space clustering within this learned manifold allowed us to group users not merely by their current value, but by their trajectory (inertia) - k = 4, cosine metric.
3. **Comparison with Traditional Methods:** Unlike traditional feature-based clustering which relies on Euclidean snapshots, our approach successfully disentangled overlapping profiles. Specifically, it resolved the ambiguity between "Active Churners" (Cluster 0) and "Vacation Crashers" (Cluster 3)—two groups that appear identical in static analysis but possess opposing velocity vectors.
4. **Visualization:** As evidenced by the UMAP and hyperspherical projections (see Appendix), the learned representations form distinct, topologically separated boundaries.

# 6 Appendix

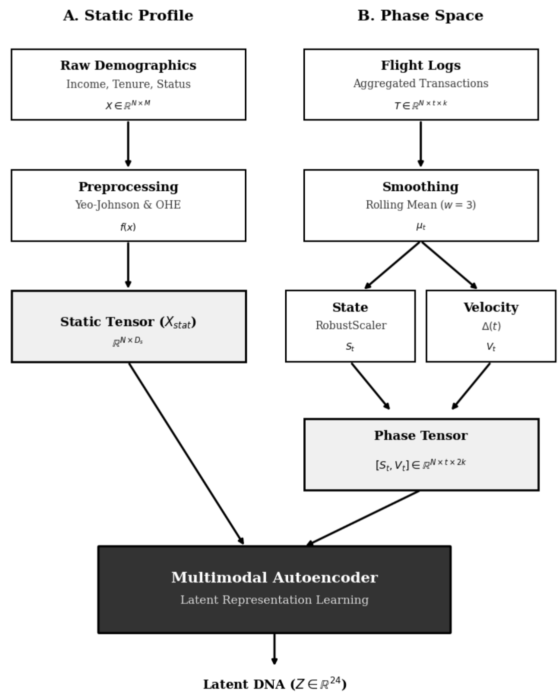


Figure 1: Data preparation composed of Static tensor (customer), Temporal tensor and delta Temporal tensor (velocity)

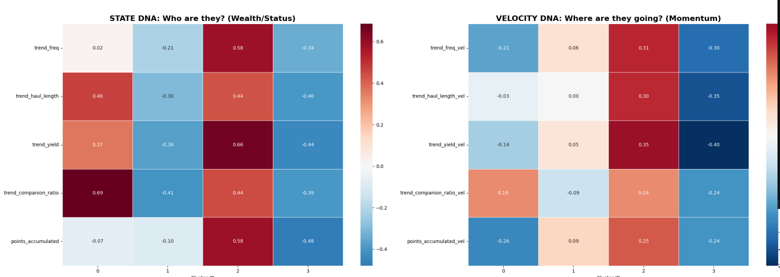


Figure 2: Customer segmentation (k=4)

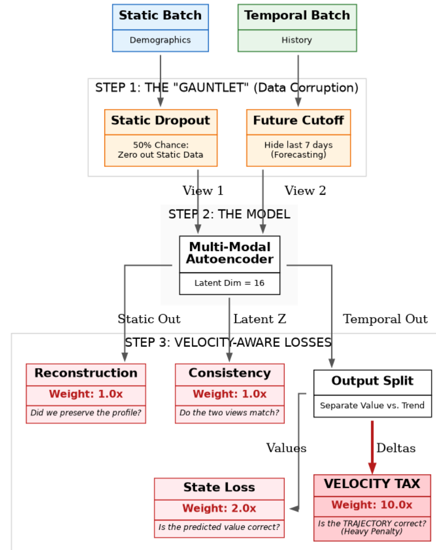


Figure 3: Training losses.

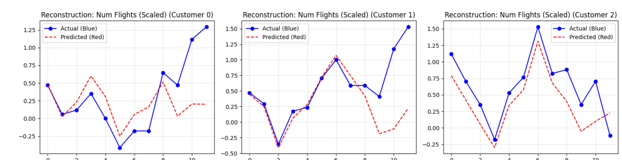


Figure 4: Temporal reconstruction forecast (num_flights)

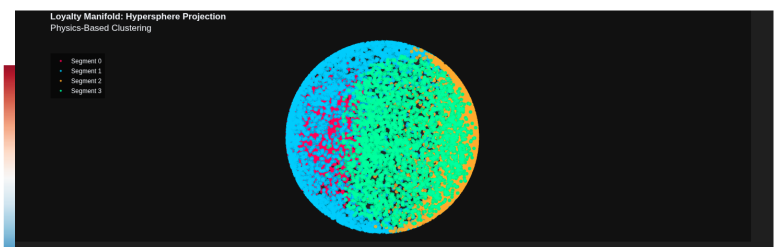


Figure 5: Clustering hypersphere representation after using PCA to spherical coordinates

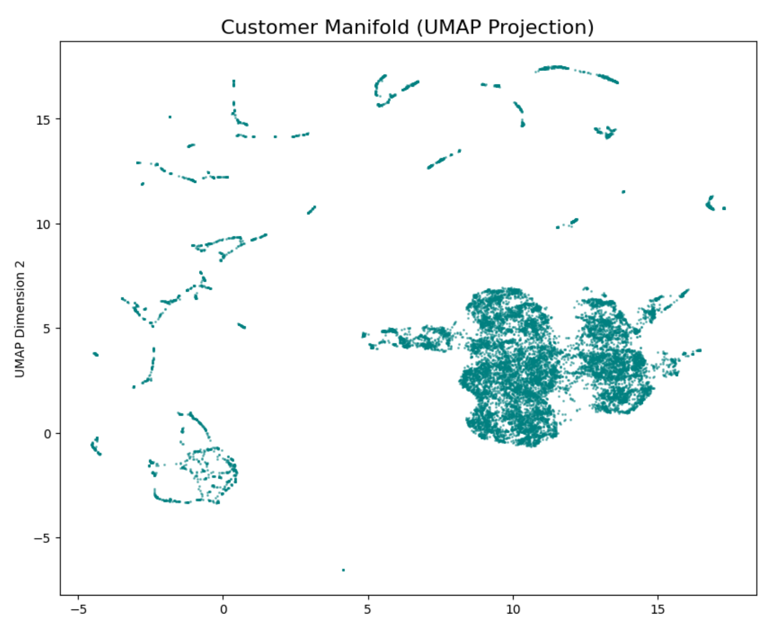


Figure 6: UMAP projection (cosine distance)