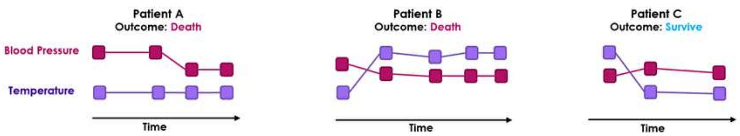
Temporal Supervised Contrastive Learning with Applications to Tabular Time Series Data

Shahriar Noroozizadeh , Jeremy C. Weiss , George H. Chen
snoroozi@andrew.cmu.edu , jeremy.weiss@nih.gov , georgechen@cmu.edu

Introduction: Supervised Contrastive Learning

Contrastive Learning:

- "Similar" data points have close embeddings
- "Dissimilar" data points have their embeddings far apart
- Classification Extension → Supervised Contrastive Learning



Temporal-SCL:

- Extending SCL to incorporate temporal dynamics
- Learning the embedding representation of time series at individual time step level

Main Contributions of Temporal-SCL

Temporal-SCL Framework Learns the embedding space with following properties:

- Predictive
- Temporally smooth
- **Diverse in capturing raw feature heterogeneity**

Clustering-based heatmap visualization relating the embedding space to both raw features and to prediction outcomes.

Experimental Findings:

- Temporal-SCL correctly recovers the underlying ground truth embedding structure of a synthetic dataset
- In two real-world clinical datasets, Temporal-SCL achieves competitive accuracy compared to various baselines

Real-world Clinical Datasets:

MIMIC Dataset (Static Outcome):

- Time series data of septic patients
- 18,354 septic patients
- Observed mortality rate ~20%
- Each time series has a single classification label

ADNI Dataset (Dynamic Outcome):

- Longitudinal data tracking the progression of Alzheimer's disease
- 11,651 hospital visits from 1,346 patients
- Different classification label at each timestep

Temporal-SCL Framework:

Temporal-SCL consists of 3 networks:

- Encoder network (f)
  Map each time step features (snapshots) to an embedding
- Predictor network (g)
  Maps the embeddings predicted class probabilities
- Temporal network (h)
  Encouraging temporal smoothness of embeddings

Jointly training the encoder and temporal networks :

$$L_{\text{overall}} = L_{\text{SCL-snapshots}} + \alpha L_{\text{temp-reg}}$$

- Supervised contrastive loss:

$$L_{\text{SCL-snapshots}} := -\sum_{((\mathbf{x}_i^{(\ell)}, y_i^{(\ell)}), (\mathbf{x}_{i'}^{(\ell')}, y_{i'}^{(\ell')})) \in \mathcal{E}_{\text{batch}}} \log \left[\frac{\exp(\langle f(\mathbf{x}_i^{(\ell)}), f(\mathbf{x}_{i'}^{(\ell')}) \rangle / \tau)}{\sum_{\substack{(\mathbf{x}_{i''}^{(\ell'')}, y_{i''}^{(\ell'')}) \in \mathcal{V}_{\text{batch}} \\ \text{s.t. } (i'', \ell'') \neq (i, \ell)}} \exp(\langle f(\mathbf{x}_i^{(\ell)}), f(\mathbf{x}_{i''}^{(\ell'')}) \rangle / \tau)} \right]$$

- Temporal-Smoothness loss:

$$L_{\text{temp-reg}} := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{L_i - 1} \sum_{\ell=1}^{L_i - 1} \left\| h((\mathbf{z}_i^{(1)}, \delta_i^{(1)}), \ldots, (\mathbf{z}_i^{(\ell)}, \delta_i^{(\ell)})) - \mathbf{z}_i^{(\ell+1)} \right\|^2$$

Finding "Similar" Pairs for Temporal-SCL:

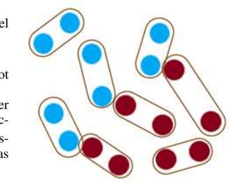
Two snapshots are considered "similar" when:

- Share the same classification outcome
- Raw feature vectors are "close to each other"

Sampling approach to find pairs of "similar: snapshots based on nearest neighbor pairing

Nearest neighbor pairing mechanism.

1. Initialize the set of snapshot pairs to be empty: $\mathcal{E} \leftarrow \emptyset$.
2. For each class $c \in [C]$:
   (a) Let the set $\mathcal{A}_c$ consist of all snapshots whose label is c.
   (b) While $|\mathcal{A}_c| \geq 2$:
      i. Let $(\mathbf{x}_i^{(\ell)}, y_i^{(\ell)})$ be a randomly chosen snapshot from $\mathcal{A}_c$.
      ii. (Nearest neighbor search) Among all the other snapshots in $\mathcal{A}_c$, find the one whose feature vector is closest to $\mathbf{x}_i^{(\ell)}$ (e.g., using Euclidean distance). Denote the resulting snapshot found as $(\mathbf{x}_{i'}^{(\ell')}, y_{i'}^{(\ell')})$.
      iii. Add the snapshot pair $((\mathbf{x}_i^{(\ell)}, y_i^{(\ell)}), (\mathbf{x}_{i'}^{(\ell')}, y_{i'}^{(\ell')}))$ to $\mathcal{E}$.
      iv. Remove $(\mathbf{x}_i^{(\ell)}, y_i^{(\ell)})$ and $(\mathbf{x}_{i'}^{(\ell')}, y_{i'}^{(\ell')})$ from $\mathcal{A}_c$.



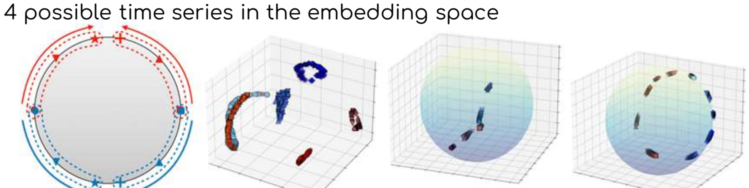
Synthetic Data Experiment:

Model	AUROC	AUPRC	Recovery
Logistic Regression	0.902±0.010	0.900±0.003	0/10
LSTM	0.951±0.008	0.948±0.002	0/10
RETAIN	0.951±0.008	0.948±0.002	2/10
DIPOLE	0.951±0.008	0.948±0.002	3/10
AC-TPC	0.951±0.008	0.948±0.002	0/10
Transformer:BERT	0.951±0.008	0.948±0.002	1/10
SIMPLE-SCL	0.805±0.007	0.804±0.002	0/10
TEMPORAL-SCL (no pretrain)	0.950±0.009	0.944±0.004	2/10
TEMPORAL-SCL (no NN pairing)	0.951±0.008	0.948±0.002	0/10
TEMPORAL-SCL (full)	0.951±0.008	0.948±0.002	**10/10**

Known ground-truth embedding structure:

- The points are all on a 2D circle
- 4 possible time series in the embedding space

(i) "● → ▼ → ★" (iii) "● → ▼ → ★"
(ii) "■ → ▲ → +" (iv) "■ → ▲ → +"



(a) Ground truth (b) Transformer (c) TEMPORAL-SCL (no NN pairing) (d) TEMPORAL-SCL

Real-World Clinical Data Experiment:

Model	MIMIC dataset		ADNI dataset	
	AUROC	AUPRC	AUROC	AUPRC
Logistic Regression	0.745±0.003	0.499±0.008	0.845±0.006	0.676±0.009
LSTM	0.767±0.003	0.509±0.005	0.947±0.002	0.823±0.005
RETAIN	0.730±0.010	0.431±0.006	0.884±0.012	0.795±0.016
DIPOLE	0.767±0.004	0.453±0.003	0.958±0.006	0.824±0.009
AC-TPC	0.703±0.006	0.432±0.007	0.839±0.013	0.681±0.017
Transformer:BERT	**0.769±0.005**	0.509±0.003	0.959±0.002	**0.922±0.003**
SIMPLE-SCL	0.744±0.003	0.486±0.003	0.902±0.024	0.796±0.020
TEMPORAL-SCL (no pretrain)	0.725±0.042	0.471±0.001	0.867±0.035	0.766±0.050
TEMPORAL-SCL (no NN pairing)	0.767±0.005	0.509±0.003	0.894±0.062	0.807±0.045
TEMPORAL-SCL (full)	0.763±0.001	**0.510±0.002**	**0.961±0.001**	0.867±0.006

Conclusion:

- Presented an extension of supervised contrastive learning to handle temporal dynamics.
- Utilized nearest neighbor pairing to capture raw feature heterogeneity
- Our pairing method alleviates the need for data augmentation in tabular time series data

Future Work: Investigate when and why nearest neighbor pairing works in different applications .

Related Works:

★ SCL *(Khosla et al. 2020)*
★ SimCLR *(Chen et al., 2020)*
★ AC-TPC *(Lee and van der Schaar, 2020)*
★ BERT-based Transformer *(Devlin et al., 2019)*
★ Dynamic-DeepHit *(Lee et al., 2019)*
★ DIPOLE *(Ma et al., 2017)*
★ RETAIN *(Choi et al., 2016)*
★ TCLR *(Dave et al., 2022)*