INFORMATICS PROFESSIONALS. LEADING THE WAY.
AMIA 2016 Annual Symposium

VAHC - Workshop on Visual Analytics in Healthcare

m-TSNE: A Framework for Visualizing High-Dimensional Multivariate Time Series

Minh Nguyen, Sanjay Purushotham, Hien To, and Cyrus Shahabi

{minhnngu, spurusho, hto, shahabi}@usc.edu

Nov. 12, 2016, Chicago, IL, USA

USC Viterbi
School of Engineering
Integrated Media Systems Center

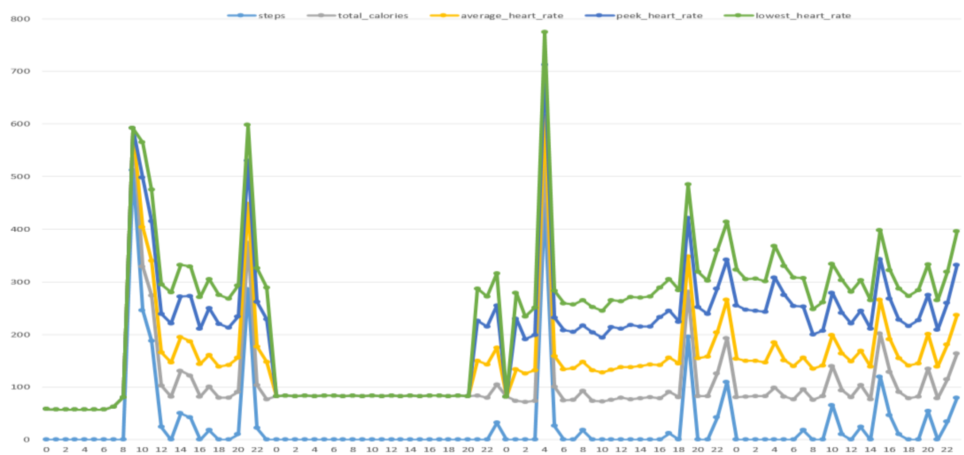
Outline

- Motivation
- Related Work
  - Univariate Time Series
  - Multidimensional Data
- Our Framework: m-TSNE (**M**ultivariate Time Series **t**-Distributed **S**tochastic **N**eighbor **E**mbedding)
- Experiment & Results
  - Human Performance ATOM-HP dataset
  - Electroencephalography EEG dataset
- Summary

Motivation

- Sensors development, e-Health platforms (EHR, mobile health,…)

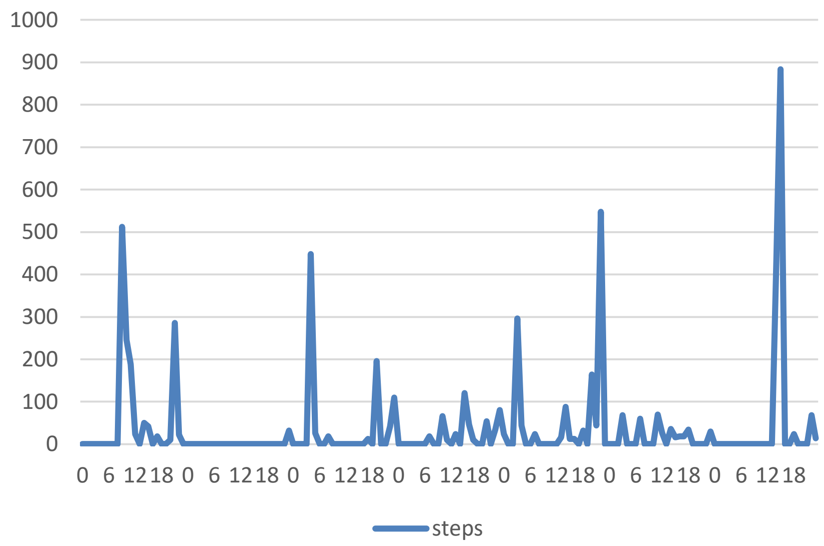
- Multivariate Time Series (MTS) in Healthcare:
  – Human vital signs
  – Remote patient monitoring data
  – Medical sensors data
  – Lab results
  – …



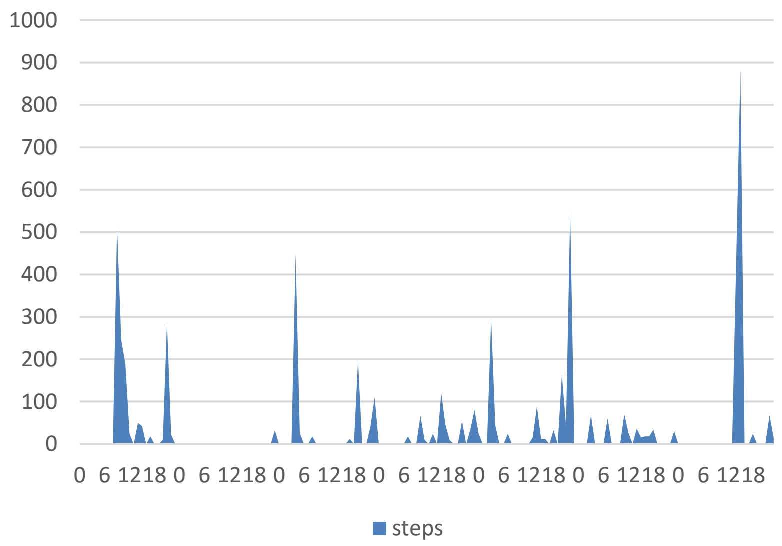
USC Viterbi
School of Engineering
Integrated Media Systems Center

Related Work

- Univariate Time Series Data:
  - Jointly visualize multiple lines / stacked graphs of features



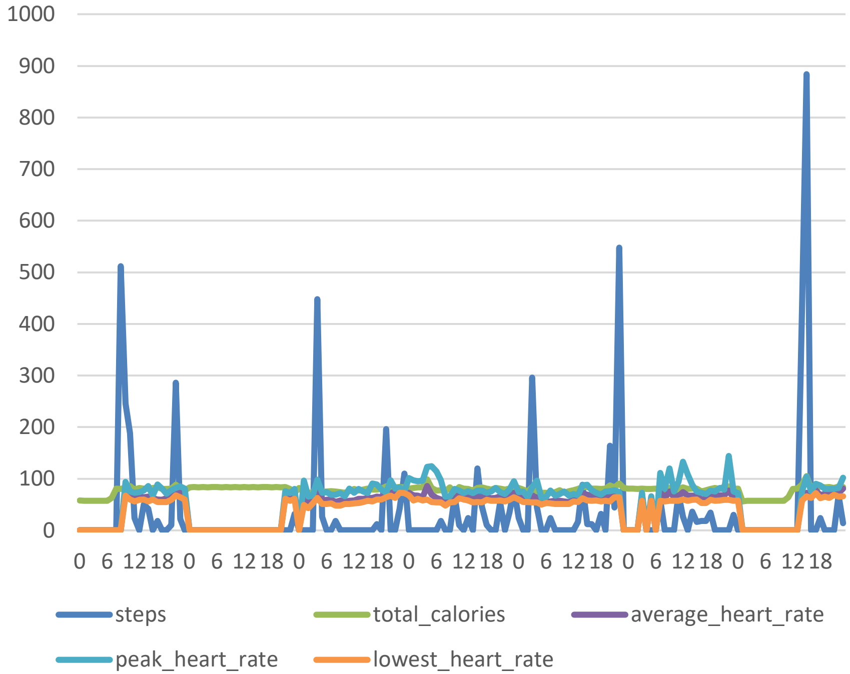
Multiple lines *[Playfair 1786]*



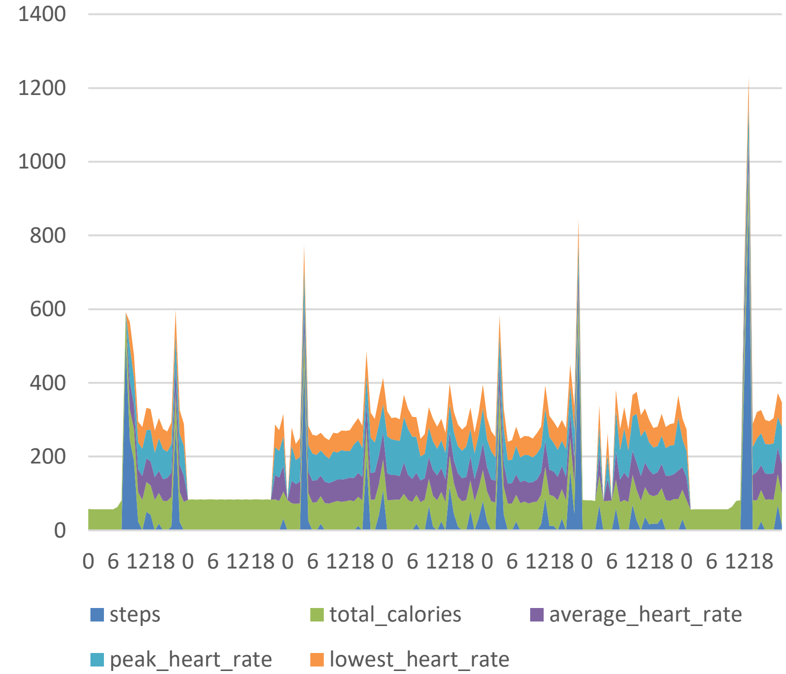
Stacked graphs *[Byron & Wattenberg '08]*

Related Work

- Univariate Time Series Data:
  - Jointly visualize multiple lines / stacked graphs of features



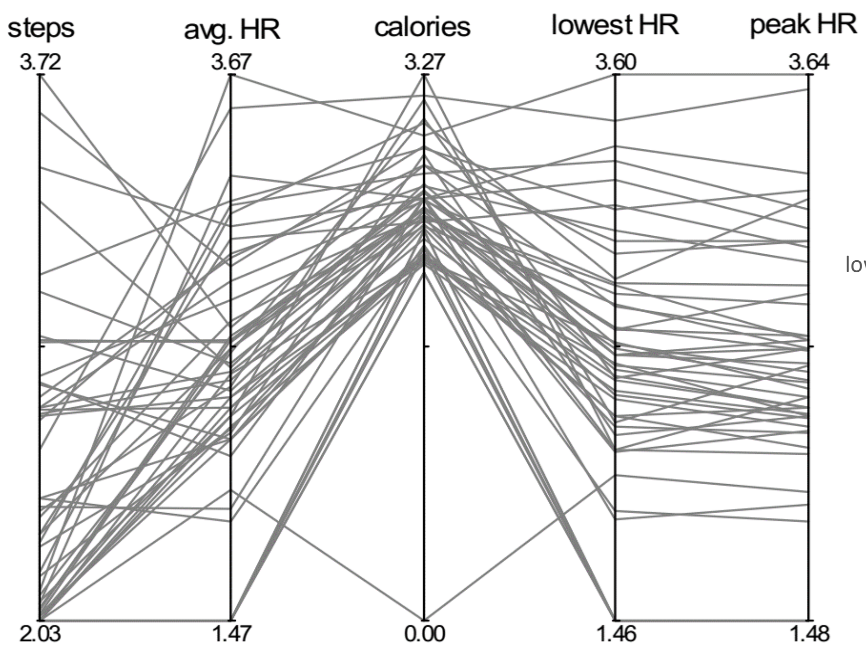
Multiple lines *[Playfair 1786]*

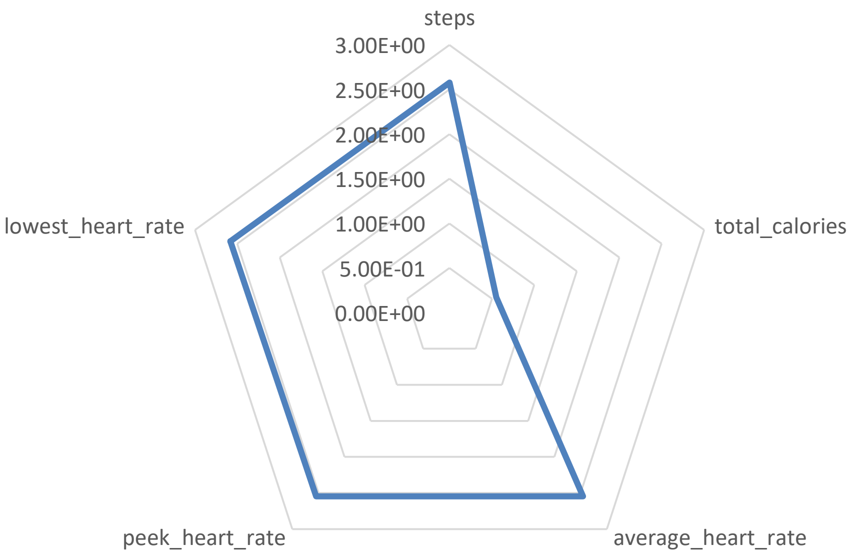


Stacked graphs *[Byron & Wattenberg '08]*

Related Work

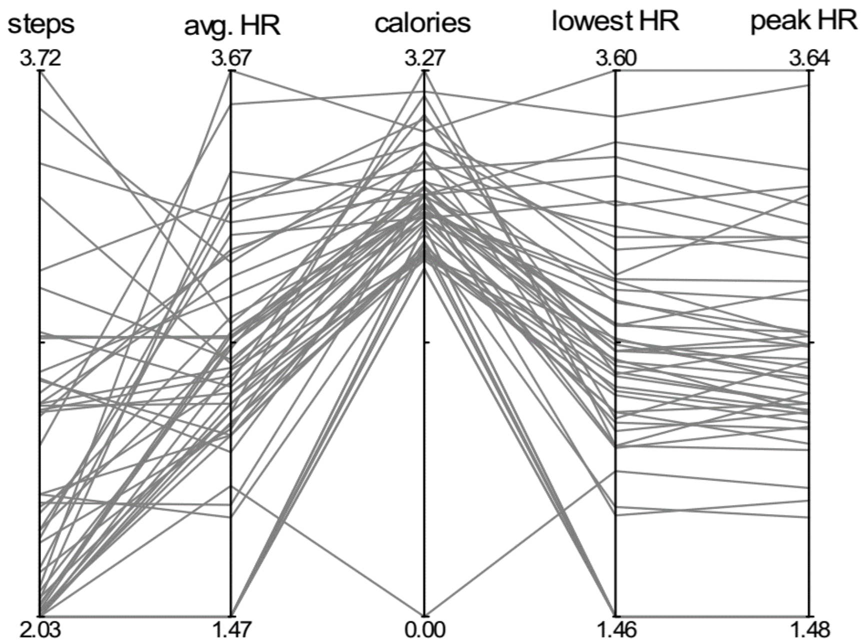
- Multidimensional Data
  - Use Radar chart *[Chambers '83]*, Parallel coordinates *[Inselberg '85]* approach

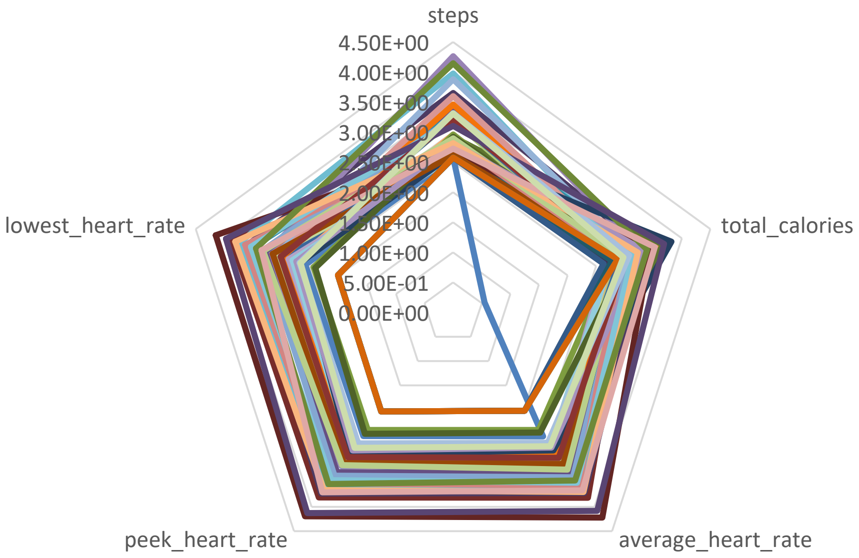




Related Work

- Multidimensional Data
  - Use Radar chart *[Chambers '83]*, Parallel coordinates *[Inselberg '85]* approach





Our Framework: m-TSNE

Challenges:

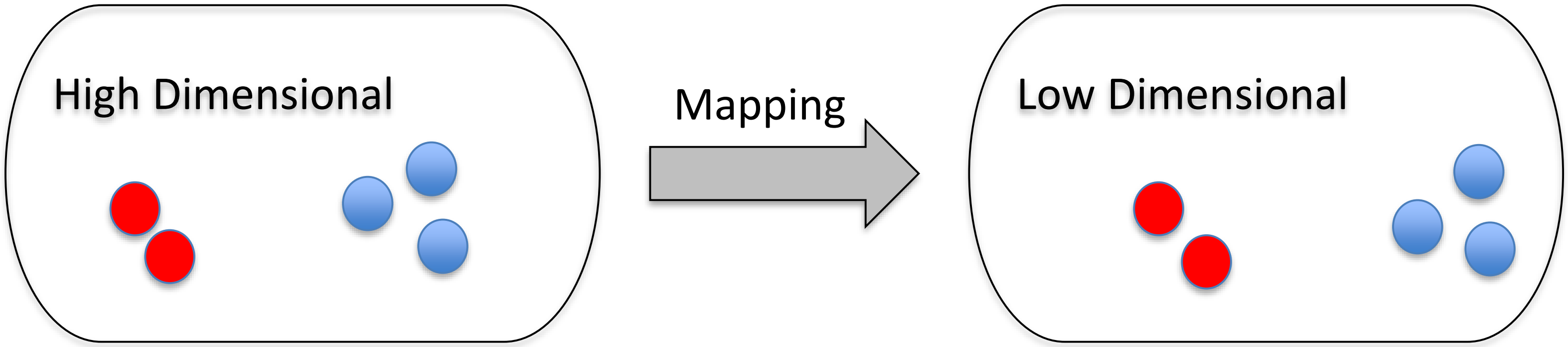
- Long time series
- Multiple dimensions
- Comparing multiple Multivariate Time Series (MTS)

→ Previous techniques focus on Refine & Represent data visually.

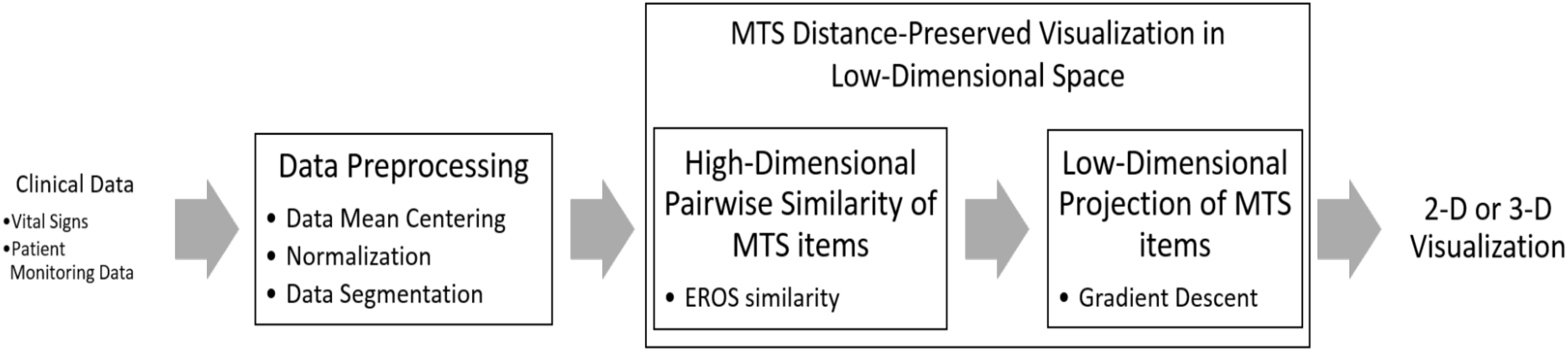
→ Using Machine Learning techniques in visualization to help / support in Data Insight Observation.

Our Framework: m-TSNE

- Consider each MTS as a data point
- Build map where points distances describe MTS similarities
- ***Embedding***: Minimize the discrepancy between *high-dimensional space* MTS data points and *low-dimensional space* data points



Our Framework: m-TSNE



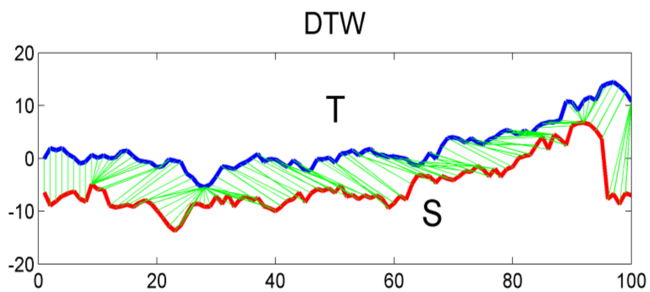
Raw data MTS X = {X_1, X_2,…, X_n}
Each X_i is a feature (a univariate time series).

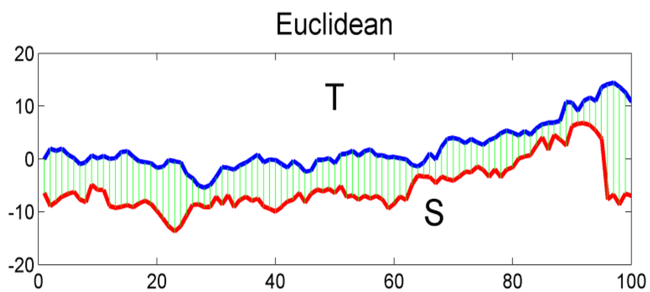
steps	total_calories	average_heart_rate	peak_heart_rate	lowest_heart_rate
246	83	75	94	67
188	86	66	75	60
24	79	63	73	56
0	82	65	74	59
50	81	64	77	60
42	80	65	86	56
0	82	62	67	60

Similarity of 2 MTS

- Each window/segmentation is an MTS X’ = X = {X'_1, X'_2,…, X'_n}. X'_i is a univariate time series within a window length (e.g. a day, a month)

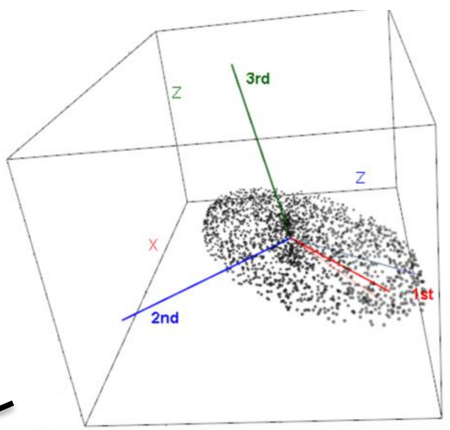
- Different similarity metrics:

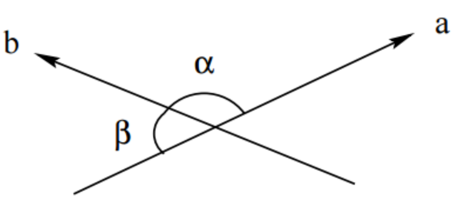




EROS: Extended Frobenius norm
[*Yang & Shahabi ‘04*]

1.756498	0.364586	1.803076	1.89864	1.68528
1.183661	1.549491	0.073897	-0.56959	0.104244
-0.43609	-1.21529	-0.5025	-0.82941	-0.7992
-0.67312	-0.03038	-0.11823	-0.6995	-0.12162
-0.1793	-0.42535	-0.31037	-0.30978	0.104244
-0.25831	-0.82032	-0.11823	0.859384	-0.7992
-0.67312	-0.03038	-0.69463	-1.60885	0.104244
-0.49534	0.364586	-1.07889	1.249105	-1.02507
-0.67312	-0.82032	-1.27102	0.20985	-1.02507
-0.67312	-0.82032	-0.69463	-1.08922	-1.02507

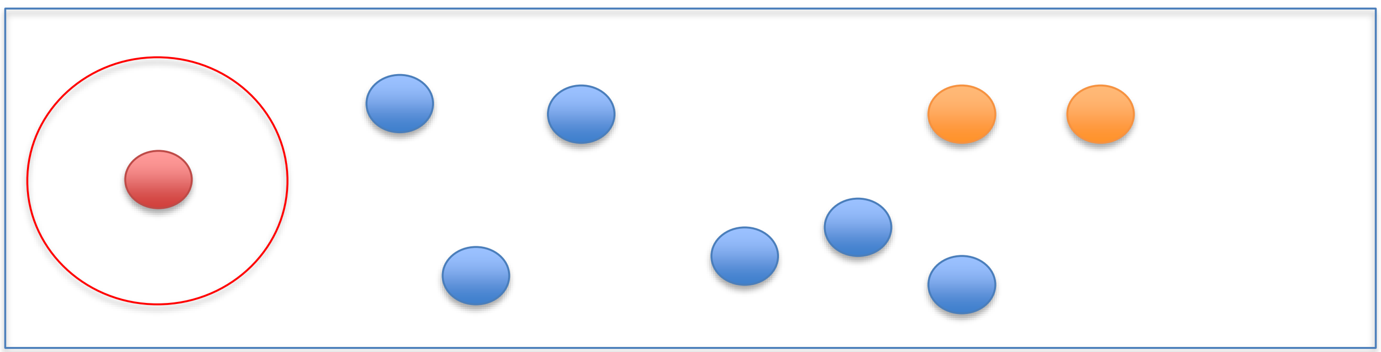




Low-dimensional Projection (t-SNE)

- **p_{ij}** : high-dimensional similarity of 2 MTS data points x_i and x_j
- **q_{ij}** : low-dimensional similarity of 2 MTS data points y_i and y_j
- Move points using gradient descent $$\frac{\delta C}{\delta y_i} = 4\sum_j (p_{ij} - q_{ij})(y_i - y_j)$$

Low Dimensional



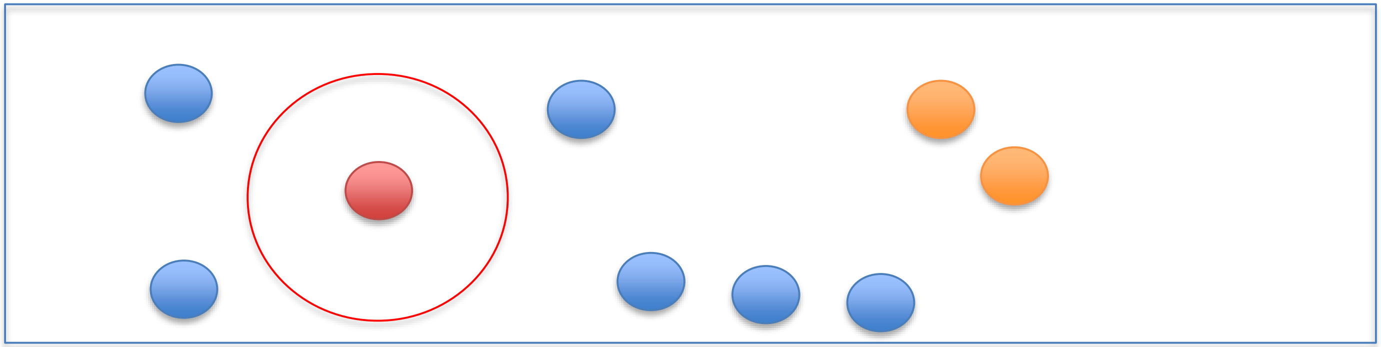
- Minimize Kullback-Leibler divergence: *[Maaten & Hinton '08]*

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Low-dimensional Projection (t-SNE)

- **p_{ij}** : high-dimensional similarity of 2 MTS data points x_i and x_j
- **q_{ij}** : low-dimensional similarity of 2 MTS data points y_i and y_j
- Move points using gradient descent $\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)$

Low Dimensional



- Minimize Kullback-Leibler divergence: *[Maaten & Hinton '08]*

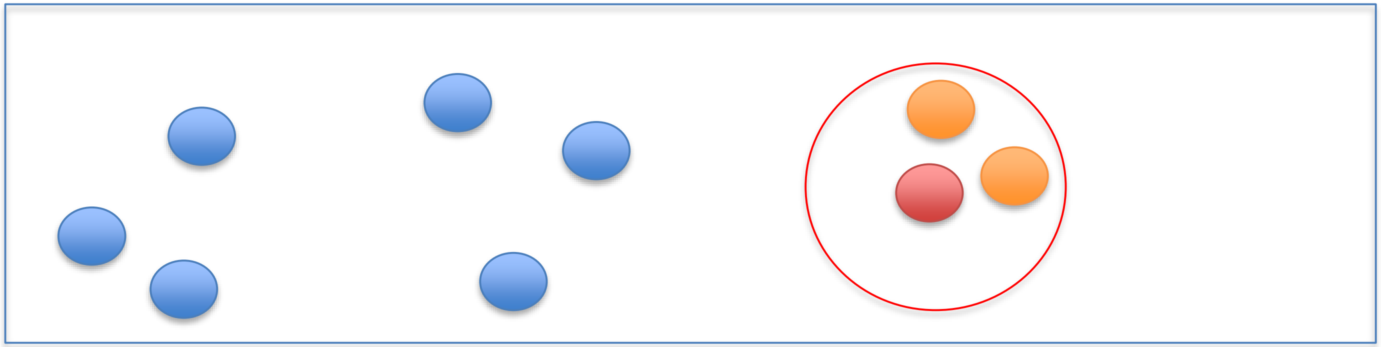
$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

USC Viterbi
School of Engineering
Integrated Media Systems Center

Low-dimensional Projection (t-SNE)

- **p_{ij}** : high-dimensional similarity of 2 MTS data points x_i and x_j
- **q_{ij}** : low-dimensional similarity of 2 MTS data points y_i and y_j
- Move points using gradient descent $\frac{\delta C}{\delta y_i} = 4\sum_j (p_{ij} - q_{ij})(y_i - y_j)$

Low Dimensional



- Minimize Kullback-Leibler divergence: *[Maaten & Hinton '08]*

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

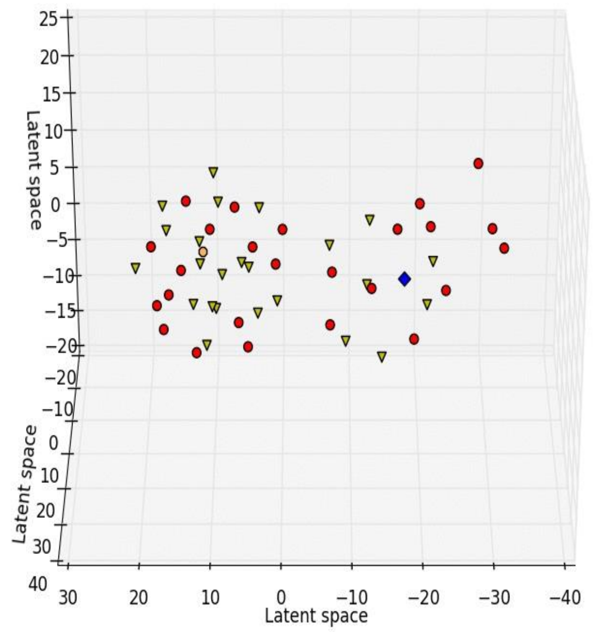
Experimental Setup

2 MTS Datasets:

Analytical Technologies to Objectively Measure Human Performance (ATOM-HP) Dataset	**Control vs. Alcoholic Electroencephalography (EEG) Dataset**
Home monitoring data of anonymized cancer patients	Control vs. alcoholic subject performing trials
5 features: Steps, Total Calories, Heartrate (average, lowest, peak)	64 features: 64 electrodes placed on the subject's scalps
2 chemotherapy cycles: 60 days	Each trial's duration is 1s.
Data sample rate: per hour	Data sample rate: 3.9-msec (256Hz)
There are 8 patients (more patients are being enrolled in this on-going study)	There are 20 subjects (10 controlled, 10 alcoholic). Each subject performs 30 trials.

Results

- ATOM-HP Dataset



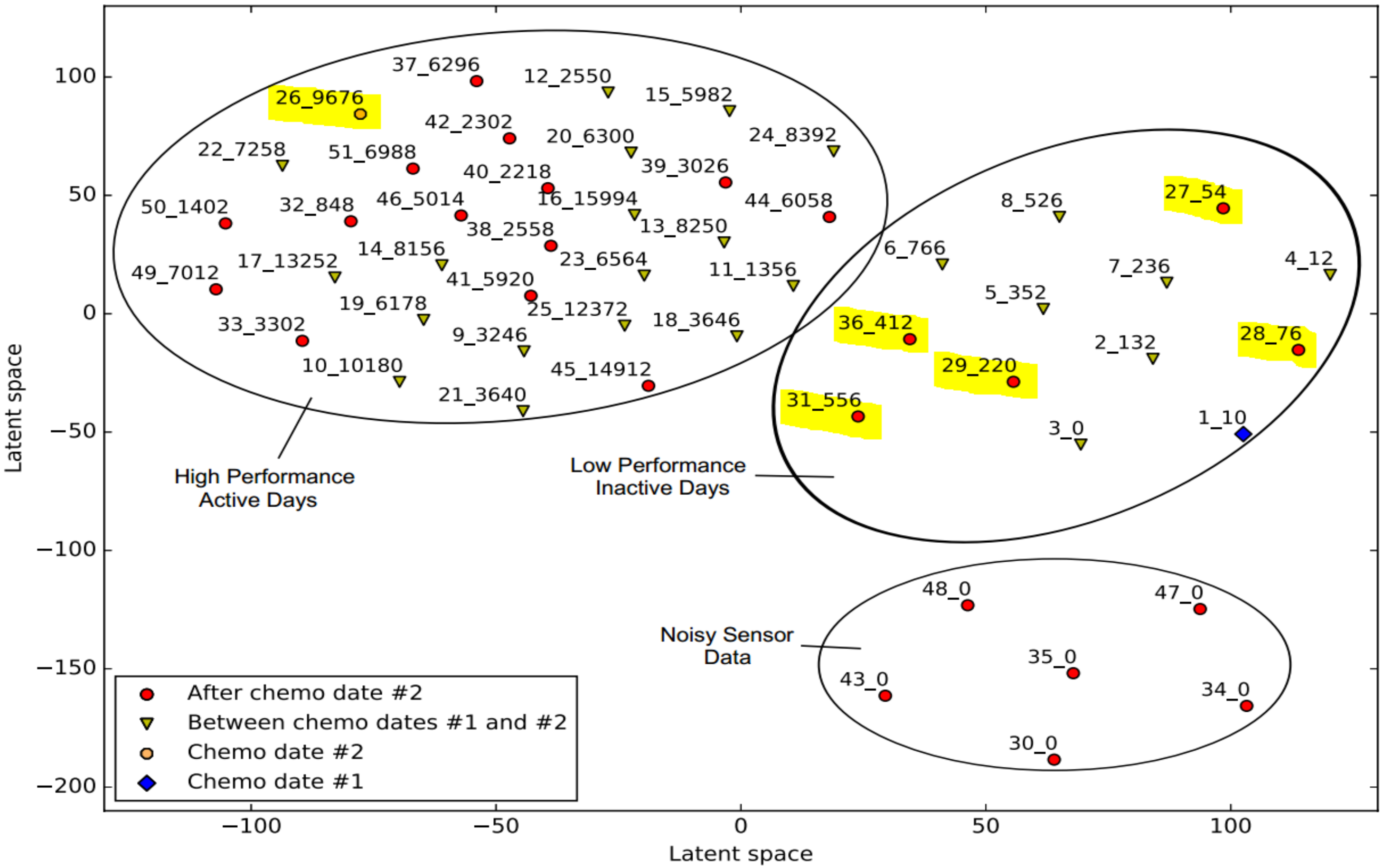
- Monitoring data of one patient
- Each point is a daily MTS

- 3 Distinct Clusters of Points:
  High Performance (Active) days
  Low Performance (Inactive) days
  Noisy Sensor Data

- Any further relationship between points and chemotherapy treatment?

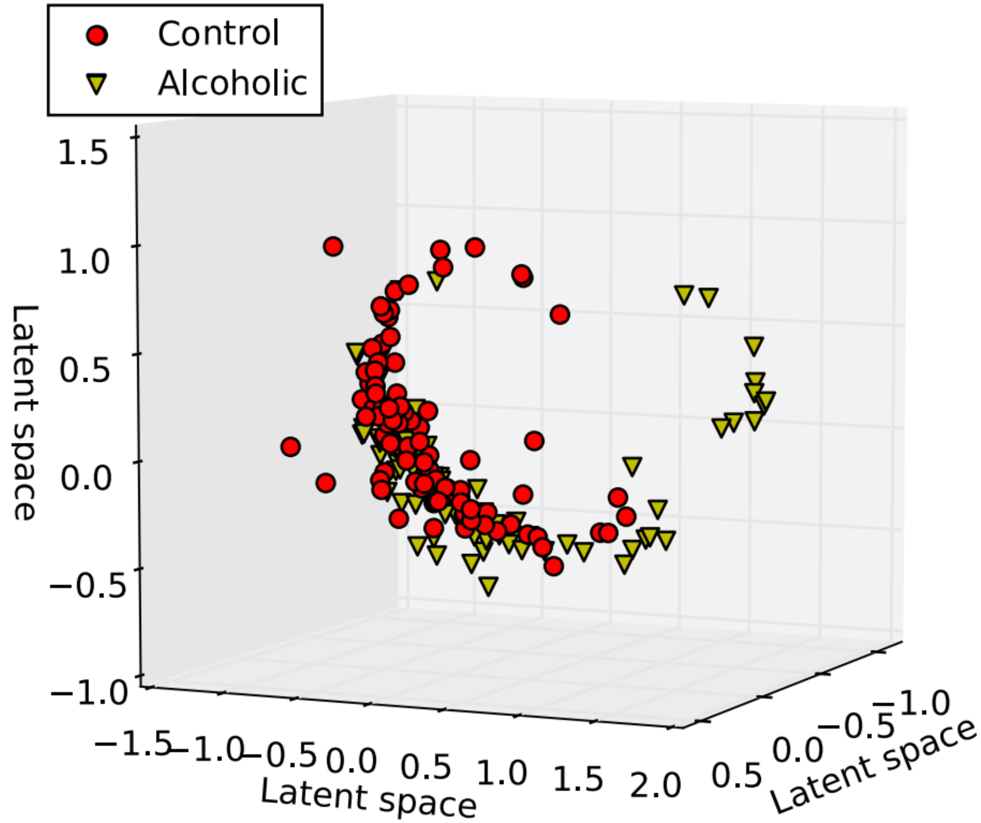
Results

- ATOM-HP Dataset



Results

- EEG Dataset



- Each point is a trial performed by a control / alcoholic subject
- Show a manifold:
  Inside: Control subject
  Outside + Outliers: Alcoholic subject

USC Viterbi
School of Engineering
Integrated Media Systems Center

Summary

- Conclusion
  - m-TSNE: a framework to visualize high-dimensional MTS data
  - Empirical evaluation on two healthcare datasets: ATOM-HP dataset, and EEG dataset

- Future Work
  - More subjects / data in on-going study ATOM-HP
  - Dynamically visualize high-dimensional MTS data
  - Adding HCI for visualization results
  - Applying using different features / variables in MTS

Q&A

Minh Nguyen
University of Southern California
----
minhnngu@usc.edu