



Tropical Cyclone Event Sequence Similarity Search via Dimensionality Reduction and Metric Learning

Shen-Shyang Ho, [Wenqing Tang](#), [W. Timothy Liu](#)

[Jet Propulsion Laboratory](#)

University of Maryland, College Park

KDD-2010

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology. It was funded by the National Aeronautics and Space Administration (NASA) Advanced Information Systems Technology (AIST) Program under grant number AIST-08-0081.



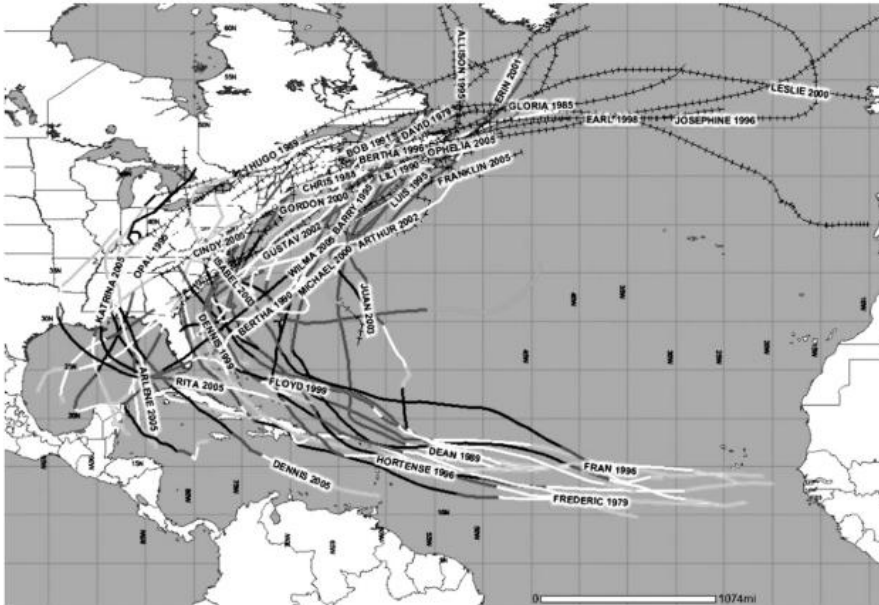
Outline

- Brief Motivation
- Methodology: Intuitions
- Some Experimental Results
- Open Problems

Brief Motivation (I)

Examples of “Conventional” similarity and objective query of interest to scientists or meteorologists:

- Track Similarity: Find all hurricanes that crossed region **A**, **B**, **C**, ...
- Intensity Similarity: Find all hurricanes that had intensity at least D km/h.
- Origin Similarity: Find all hurricanes that evolved from region **E**.



“Find all hurricanes that crossed Windsor, Ontario, and St. John's, Newfoundland from 1979 to 2005.”
Shawn M. Milrad, Eyad H. Atallah, and John R. Gyakum, “Dynamical and Precipitation Structures of Poleward-Moving Tropical Cyclones in Eastern Canada, 1979--2005”, Monthly Weather Review, vol. 137, pp. 836-851, Mar. 2009



Brief Motivation (II)

Applications Places System 8:18 AM

AIST Maps GUI - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://berlin.dbcenter.cise.ufl.edu/AIST_WEB/ Google

Most Visited Red Hat Red Hat Magazine Red Hat Network Red Hat Support

Moving Object Database GUI

for weather event analysis and tracking

Search STQL

[Back](#)

Below is a list of prepared queries

- [Find all hurricanes that have crossed Louisiana state](#) [STQL]
- [Find all US states that have been affected by the hurricane KATRINA](#) [STQL]
- [Find all states that have been crossed by all hurricanes that last longer than 10 days](#) [STQL]
- [Find US states that have never been affected by any hurricanes in 2005](#) [STQL]
- [Find all hurricanes with life time longer than 7 days](#) [STQL]
- [Find the average life time of all hurricanes](#) [STQL]
- [Find all hurricanes that have their life time overlap with hurricane OPHELIA](#) [STQL]
- [Find all hurricanes that exist during the life time of hurricane OPHELIA](#) [STQL]
- [Find all hurricanes that exist prior to hurricane OPHELIA](#) [STQL]

Map Visualization

Map Satellite Hybrid

Map data ©2010 AND, Europa Technologies, INEGI, MapLink, Tele Atlas - Terms of Use

Data Table

r.id : NUMBER	r.sname : VARCHAR2	r.geoshape : REGION2D	m.name : VARCHAR2	m.track : MPOINT
18	Louisiana	REGION2D()	CINDY	MPOINT()
18	Louisiana	REGION2D()	KATRINA	MPOINT()
18	Louisiana	REGION2D()	RITA	MPOINT()

AIST Project - UF, NASA - 2010 | Last Updated: 21 June 2010.

Done

sho@lmc-00... [MATLAB 7.8... [Command ... [Workspace] [Editor - /usr/l... sho@lmc-00... AIST Maps GU... Starting Take ...

Find all hurricanes that have crossed Louisiana state in 2005.

M. Schneider, S.-S. Ho, T. Chen, A. Khan, G. Viswanathan, W. Tang, and W. T. Liu, Moving Objects Database Technology for Ad-Hoc Querying and Satellite Data Retrieval of Dynamic Atmospheric Events, 2010 Earth Science Technology Forum, June 22-24, Arlington, VA, 2010.





Brief Motivation (III)

Applications Places System 8:26 AM

AIST Maps GUI - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://berlin.dbcenter.cise.ufl.edu/AIST_WEB/

Most Visited Red Hat Red Hat Magazine Red Hat Network Red Hat Support

Moving Object Database GUI

for weather event analysis and tracking

Search STQL

[Back](#)

Below is a list of prepared queries

- [Find all hurricanes that have crossed Louisiana state](#) [STQL]
- [Find all US states that have been affected by the hurricane KATRINA](#) [STQL]
- [Find all states that have been crossed by all hurricanes that last longer than 10 days](#) [STQL]
- [Find US states that have never been affected by any hurricanes in 2005](#) [STQL]
- [Find all hurricanes with life time longer than 7 days](#) [STQL]
- [Find the average life time of all hurricanes](#) [STQL]
- [Find all hurricanes that have their life time overlap with hurricane OPHELIA](#) [STQL]
- [Find all hurricanes that exist during the life time of hurricane OPHELIA](#) [STQL]
- [Find all hurricanes that exist prior to hurricane OPHELIA](#) [STQL]

Map Visualization

Map Satellite Hybrid

Data Table

12 rows retrieved.

m.id : NUMBER	m.name : VARCHAR2	m.track : MPOINT	m.track.getLifeTime() : NUMBER
4	DENNIS	MPOINT()	13
5	EMILY	MPOINT()	10
6	FRANKLIN	MPOINT()	9

AIST Project - UF, NASA - 2010 | Last Updated: 21 June 2010.

javascript: executePreQuery(4);

sho@lmc-00... [MATLAB 7.8... [Command ... [Workspace] [Editor - /usr/l... sho@lmc-00... AIST Maps GU... Starting Take ...

Find all hurricanes that have life-time longer than 7 days in 2005



Brief Motivation (IV)

Query with instance-level constraints/user-defined “subjective” data constraints

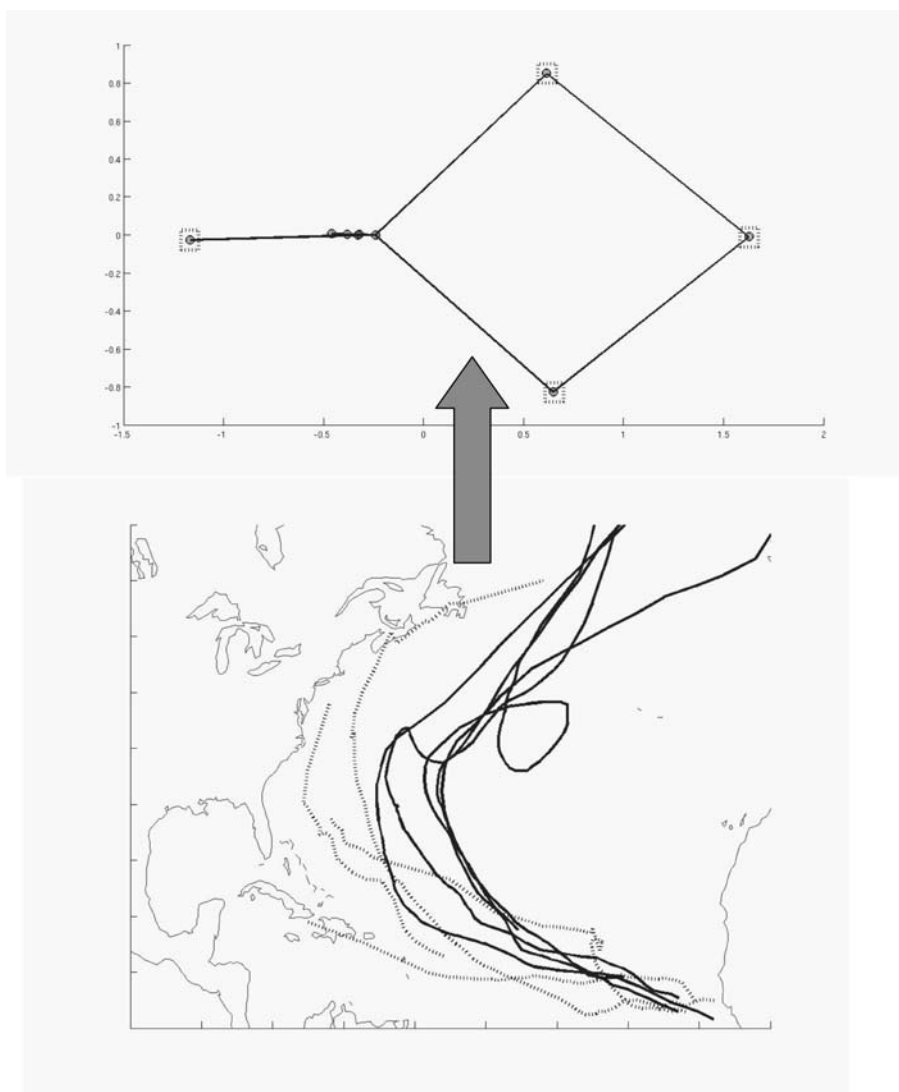
Find all hurricanes that are *similar* to hurricane A, B, C, ... and *dissimilar* to hurricane a, b, c, ...

General Data Assumptions:

Multidimensional (Spatial, Temporal, Features/Attributes)
Arbitrary Length Sequences



Methodology (I)





Standard Data Mining Techniques Used:

- **Similarity Metric: Longest Common Subsequence (LCSS) – [Vlachos et al., ICDE, 2002] - generalize to multidimensional sequences**

$$S1(A, B, \delta, E) = \frac{LCSS_{\delta, E}(A, B)}{\min(|A|, |B|)} \quad LCSS_{\delta, E}(A, B) = \begin{cases} 0 & \text{if } |A| = 0 \\ & \text{or } |B| = 0 \\ 1 + LCSS_{\delta, E}(Head(A), Head(B)) & \text{if } c_k > 0, \forall c_k, \\ & |t_i - t_j| < \delta \\ & \text{is satisfied} \\ \max(LCSS_{\delta, E}(Head(A), B), LCSS_{\delta, E}(A, Head(B))) & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} \epsilon_1 - |a_{1,t_i} - b_{1,t_j}| \\ \vdots \\ \epsilon_m - |a_{m,t_i} - b_{m,t_j}| \end{pmatrix}$$

- **Dimensionality Reduction: Isometric Feature Mapping (ISOMAP) – [Tenenbaum et al, Science, 2000]**
- **Metrics Learning: [Xing et al, NIPS, 2002]**

$$\min_{E, \delta} \sum_{(x_i, x_j) \in S} \|f_{S1}(x_i) - f_{S1}(x_j)\|^2$$

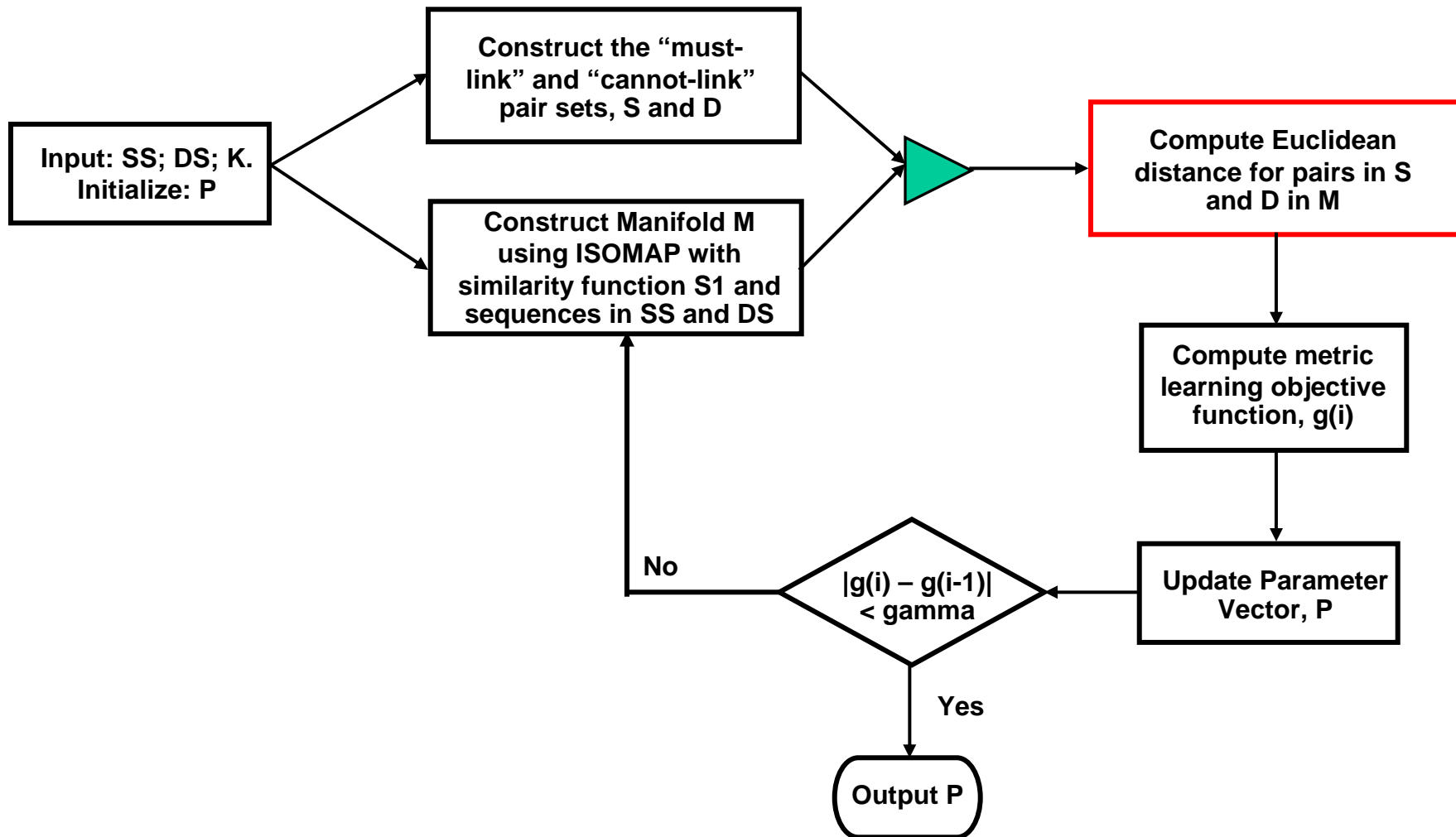
such that $\sum_{(x_i, x_j) \in D} \|f_{S1}(x_i) - f_{S1}(x_j)\| \geq 1$

and $P > 0$ where $P = (\epsilon_1, \epsilon_2, \dots, \epsilon_m, \delta) \in (R^+)^m \times Z^+$.

- **Similarity Search: Voting Approach**

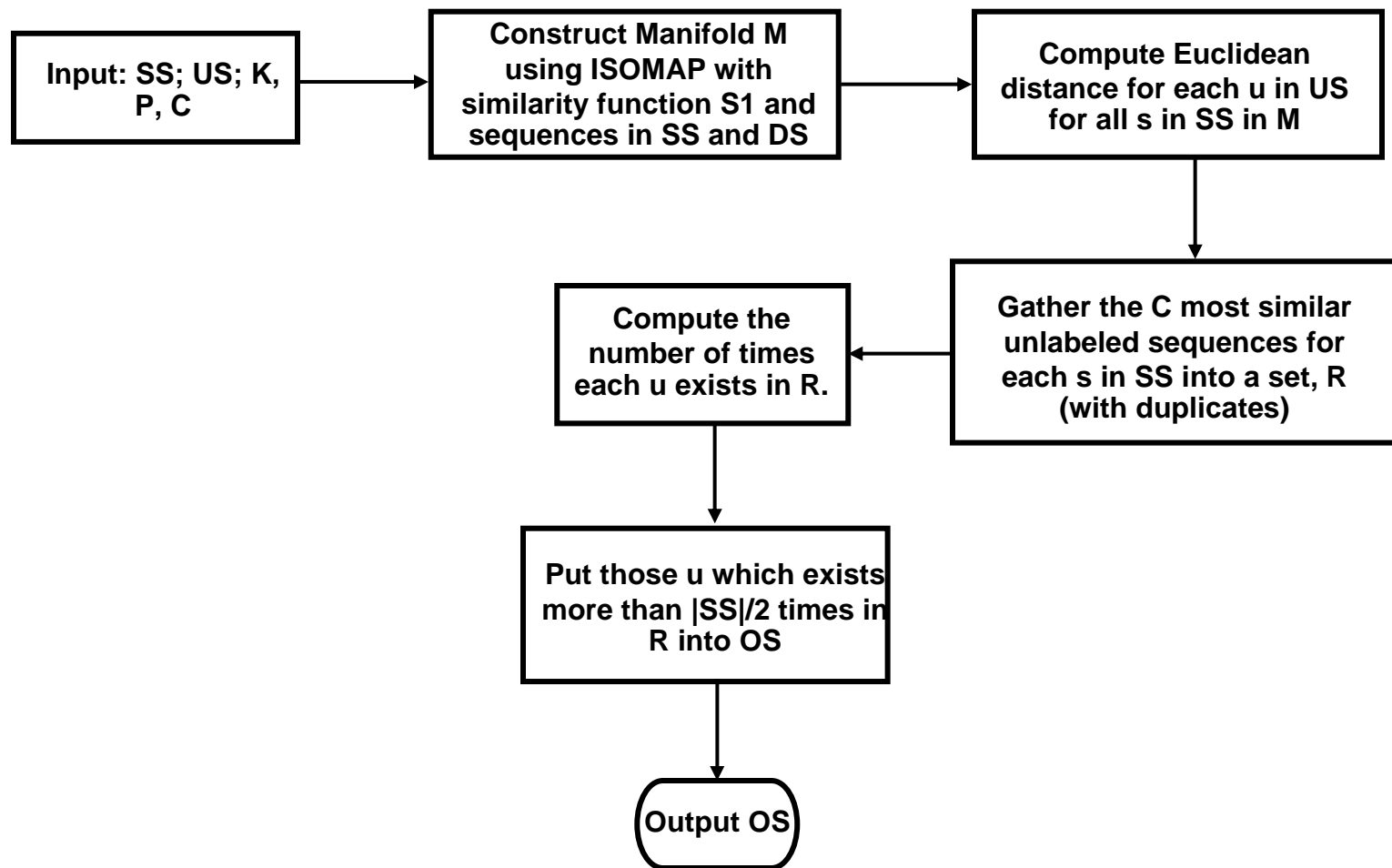


Methodology (III) – Learning LCSS Parameter P





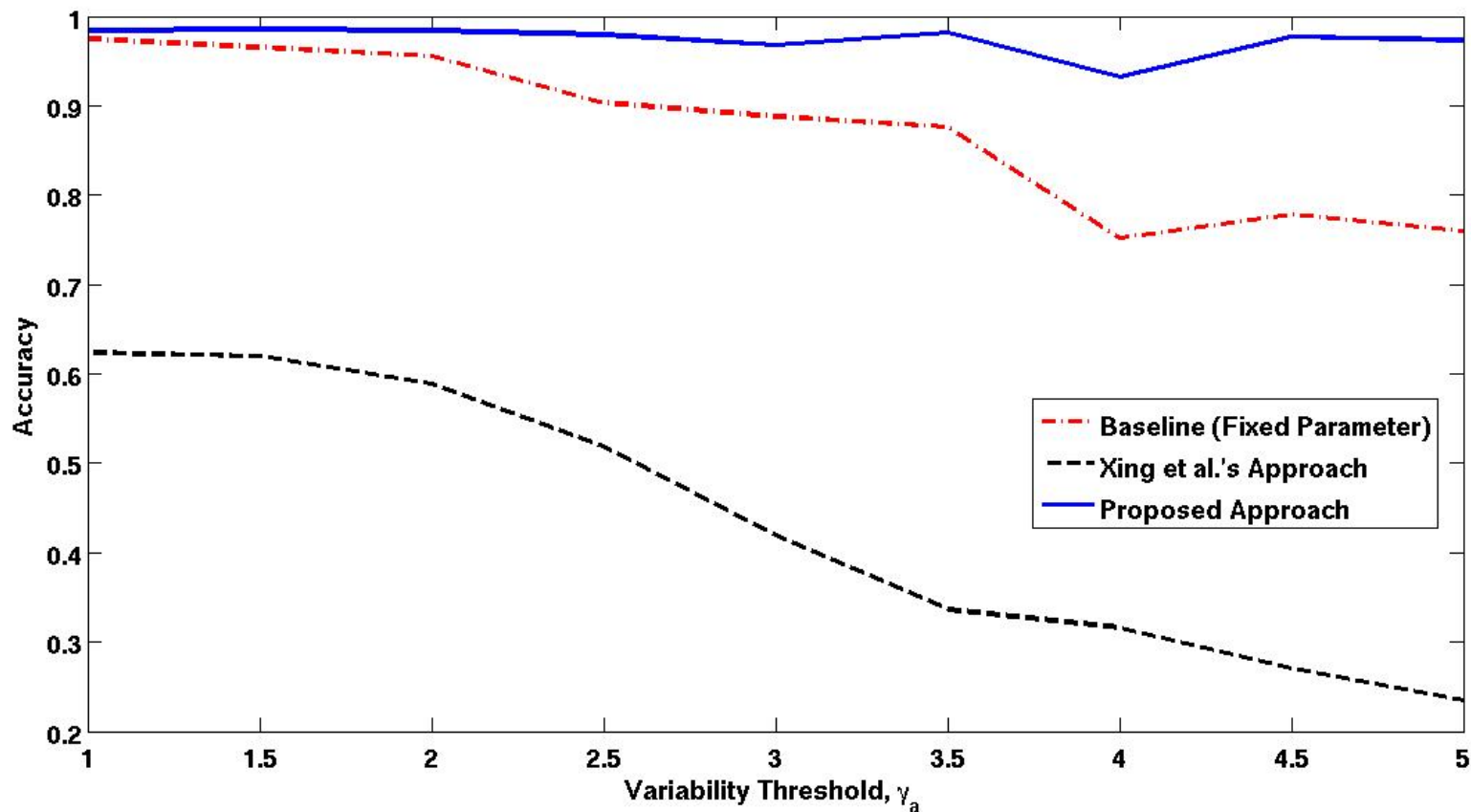
Methodology (IV) – Similarity Search on Unlabeled Sequences U'





Experimental Results (I) – Learning P

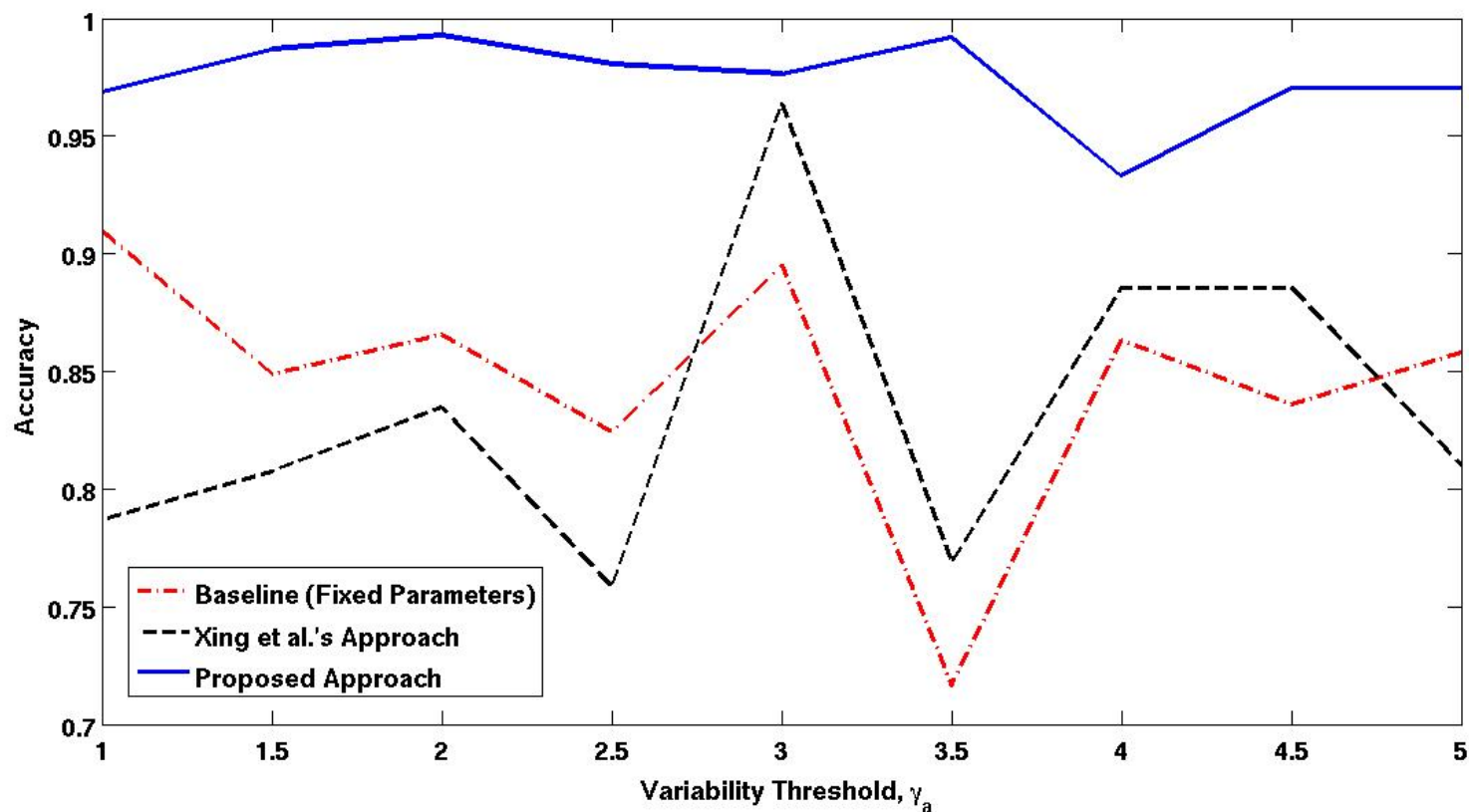
Accuracy using S1 measure in the data sequence input space





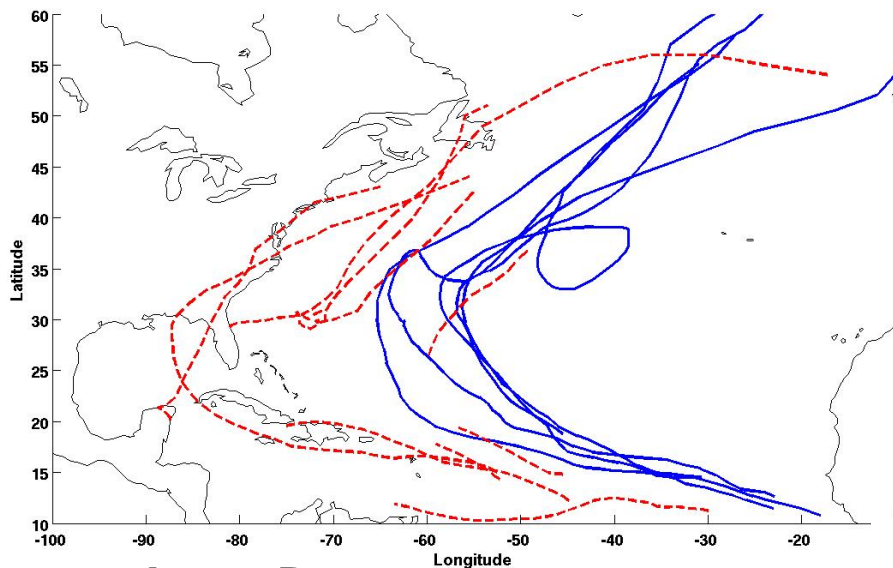
Experimental Results (II) – Learning P

Accuracy using Euclidean distance in the low dimensional manifold

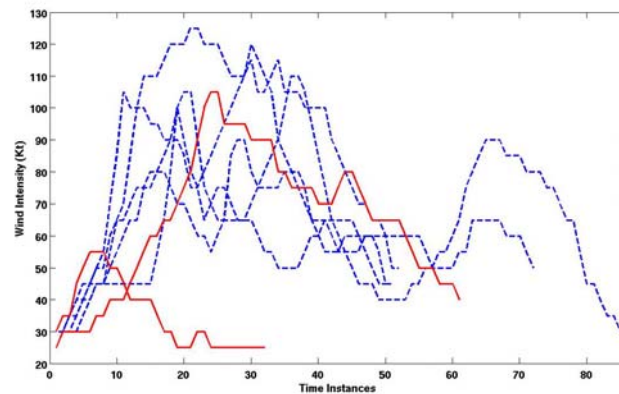




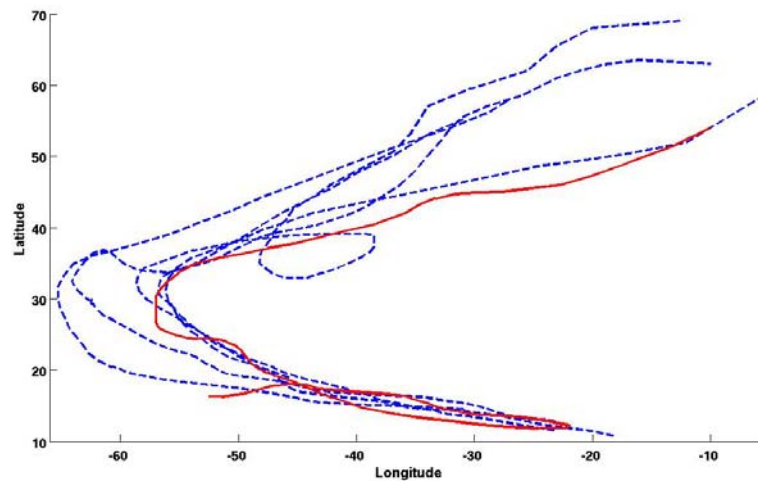
Experimental Results (III) – Similarity Search



Input Data



Output Result (Red)





Open Problems

- Theoretical justification for the setting and data assumptions.
- Which is the best similarity metric, dimensionality reduction approach, and metric learning approach for the framework?
- Local/partial sequence similarity search
- Other Applications ...

See You at the
Poster !!