Computing Correlation Anomaly Scores using Stochastic Nearest Neighbors

Tsuyoshi (Tsuyo) Idé,
IBM Research, Tokyo Research Lab.
Spiros Papadimitriou, and Michail Vlachos
IBM T.J. Watson Research Center

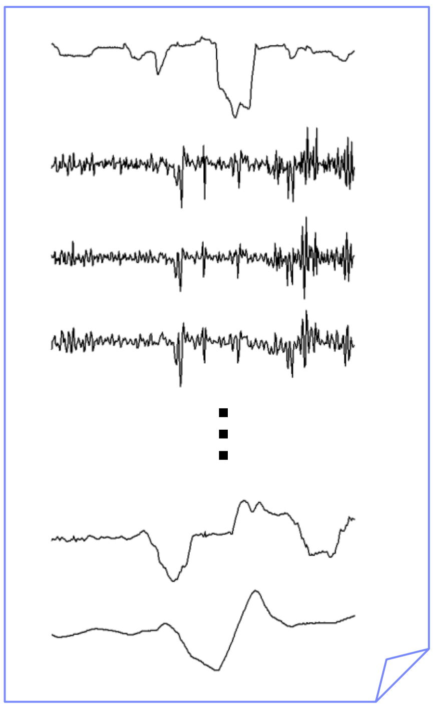
Outline

- **Problem statement**
- **Neighborhood preservation principle**
- **Stochastic nearest neighbors**
- **Experimental results and summary**

Problem statement

Problem statement (1/2):
We address a task of ***change analysis*** between two data sets

data set A



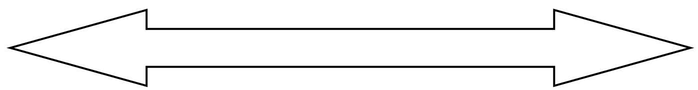
x_1

x_2

x_N

Problem 1 (change detection):

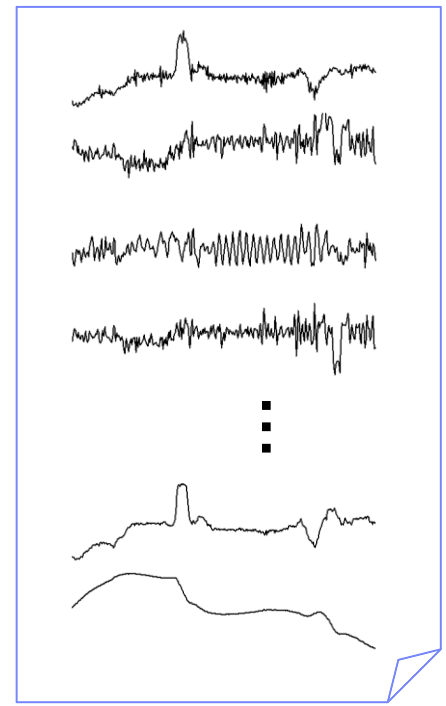
Tell whether A and B are different



Problem 2 (change ***analysis***):

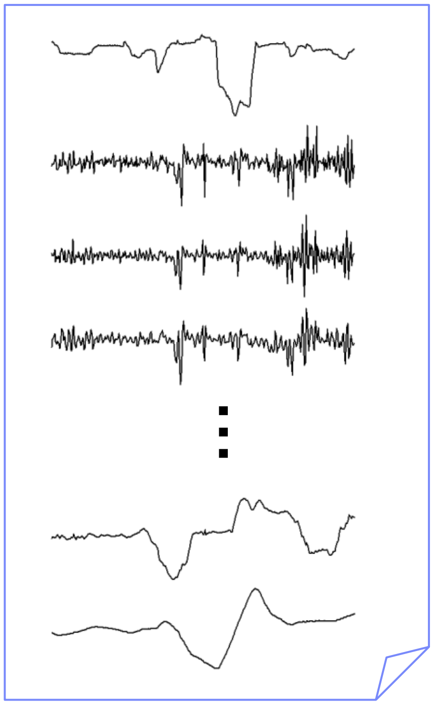
Given A and B, tell which variables are responsible for the difference between them

data set B



Problem statement (2/2): We assume sensor signals of highly correlated and dynamic natures

<u>data set A</u>



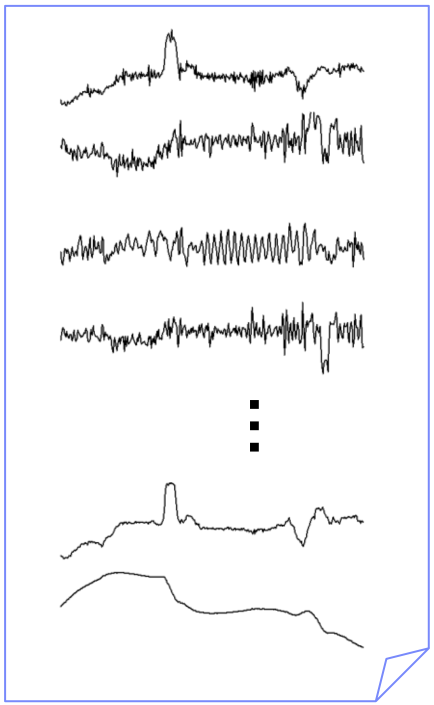
Typical application

Sensor validation (to identify faulty sensors)

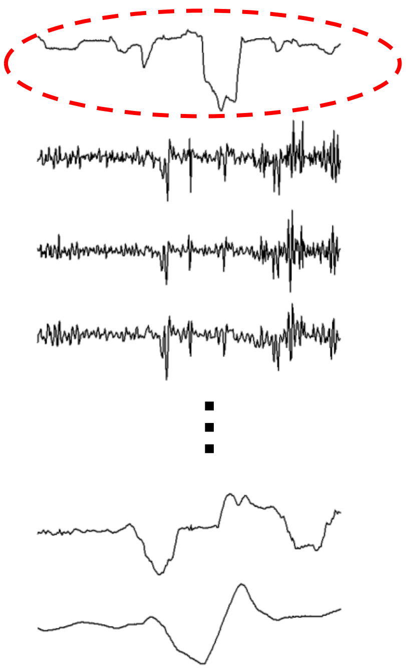
Challenges in real data

- dependency between signals
- highly dynamic nature
- heterogeneities
- no supervised information

<u>data set B</u>



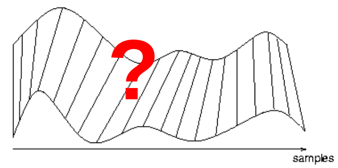
Related work:
Highly dynamic and correlated natures make the problem difficult



Time-series alignment (or DTW)

[Berndt 94, Keogh 00, …]

- hard to handle highly dynamic natures



Two-sample test

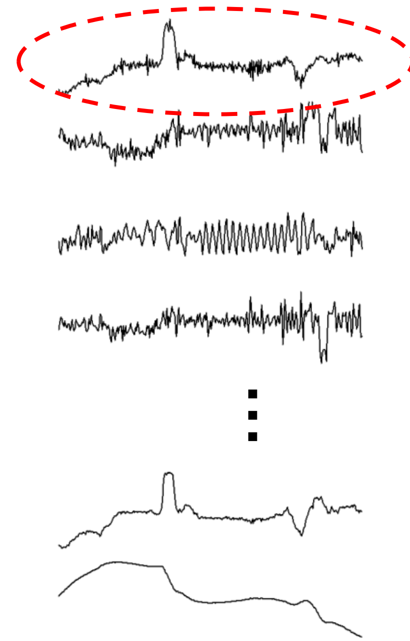
[Friedman 79, Henze 88, Gretton 07, …]

- capable of handling change detection
- but hard to do change analysis

PCA-based approach

[Papadimitriou 05, Idé 05, …]

- doesn't work since no stable latent structure in this case
  → see Experiment



Neighborhood preservation principle

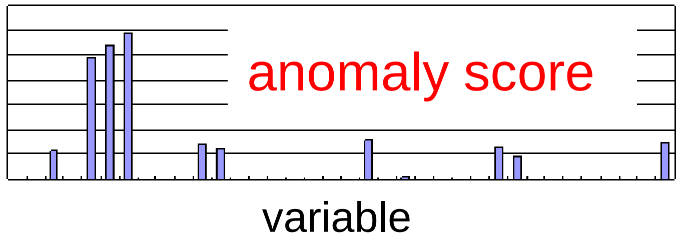
Our goal: to compute the anomaly score of each signal

data set A

test data

$$\left\{ \boldsymbol{x}^{(t)} \in \mathbb{R}^N \right\}$$

t = (time index)



data set B

reference data

$$\left\{ \bar{\boldsymbol{x}}^{(t)} \in \mathbb{R}^N \right\}$$

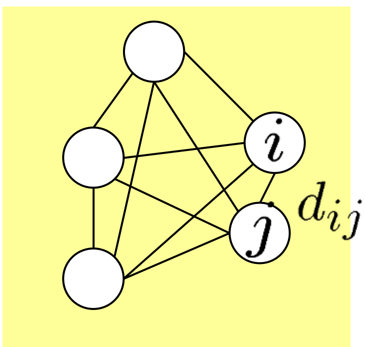
t = (time index)

Reducing the problem to graph comparison

data set A

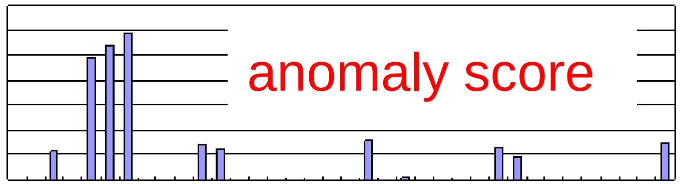
test data

dissimilarity graph



$\mathsf{D} \in \mathbb{R}^{N \times N}$

	x_1	x_2	..
x_1	0	0.2	..
x_2	0.2	0	..
..	..	..	..



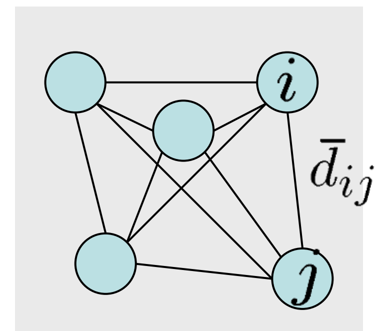
variable

Problem

Which nodes are responsible for the difference between the two graphs?

data set B

reference data



$\bar{\mathsf{D}} \in \mathbb{R}^{N \times N}$

Simplest choice of dissimilarity:

$$d_{i,j} = -\log |a_{i,j}|$$

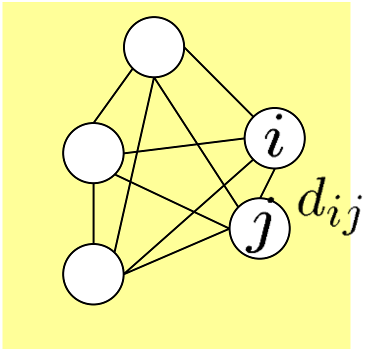
Correlation coefficient between the *i*- and *j*-th signals

Key observation: Globally unstable, but locally stable

data set A

test data

dissimilarity graph



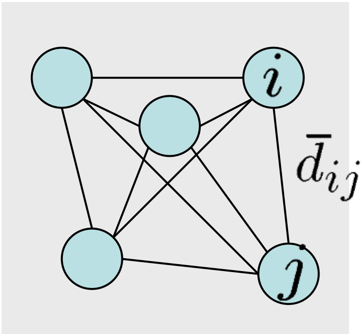
$\mathsf{D} \in \mathbb{R}^{N \times N}$

- **Global graph structure is unstable**
  - due to highly dynamic nature
- **Highly correlated pairs are relatively stable**
  - even under dynamic fluctuation

- ***Neighborhood Preservation Principle***
  - Under normal system operations, "*tightness*" of highly correlated pairs will be unchanged

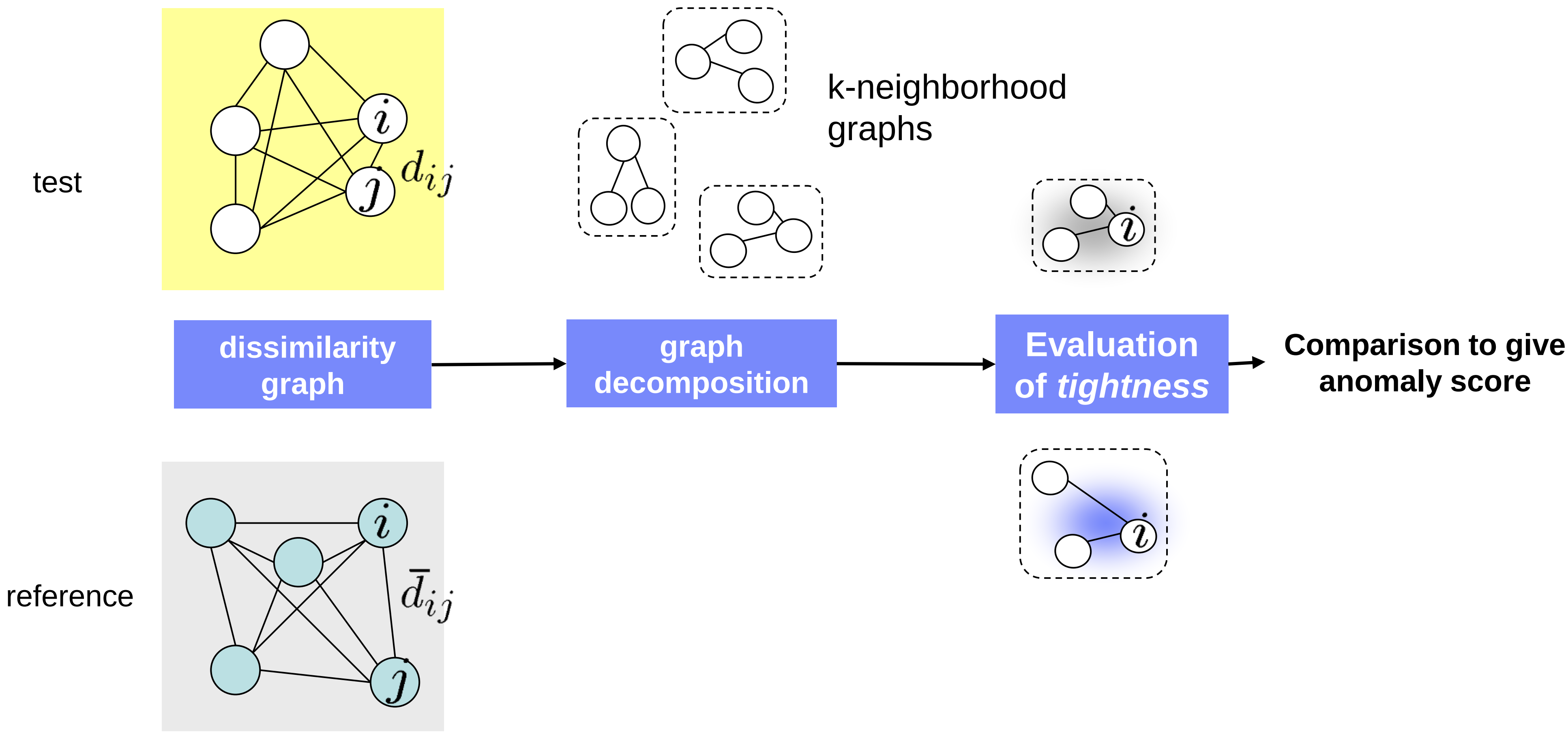
data set B

reference data



$\bar{\mathsf{D}} \in \mathbb{R}^{N \times N}$

High level overview of our approach:
We focus only on local structures of the graph



Stochastic nearest neighbors

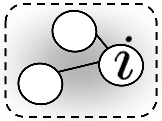
The tightness is defined as the sum of coupling probabilities $p(j|i)$

- Imagine that graph edges are not static but *stochastic*
- **The definition of tightness**

$$e_i(\mathcal{N}_i) \equiv \sum_{j \in \mathcal{N}_i} p(j|i)$$

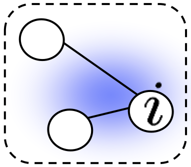
$p(j|i)$: coupling probability of i to j

$\mathcal{N}_i$: set of neighboring nodes of i



$e_i(\mathcal{N}_i)$

Evaluation of *tightness* → **Comparison to give anomaly score**



$\bar{e}_i(\mathcal{N}_i)$

- The anomaly score (***E-score***) is naturally given by*

$$E \equiv |e_i(\mathcal{N}_i) - \bar{e}_i(\mathcal{N}_i)|$$

* In fact, the algorithm has been designed to be symmetric between the two data sets. For detail, see the paper.

$p(j|i)$ is determined by utilizing a notion of stochastic neighborhood

$p(j|i)$ can be determined by solving the following problem:

→ c.f. Hinton-Roweis 03

"*For a given # of edges, minimize the average dissimilarity within the neighborhood graph*"

$$\min\langle d_i\rangle \ \text{ s.t. } \ \mathrm{e}^{H_i} = \text{const.}, \quad p(i|i) + \sum_{j\in\mathcal{N}_i} p(j|i) = 1$$

- Minimum average dissimilarity
- Constant perplexity (H_i: entropy)
  → constant # of neighboring nodes
- Normalization condition

Solution:

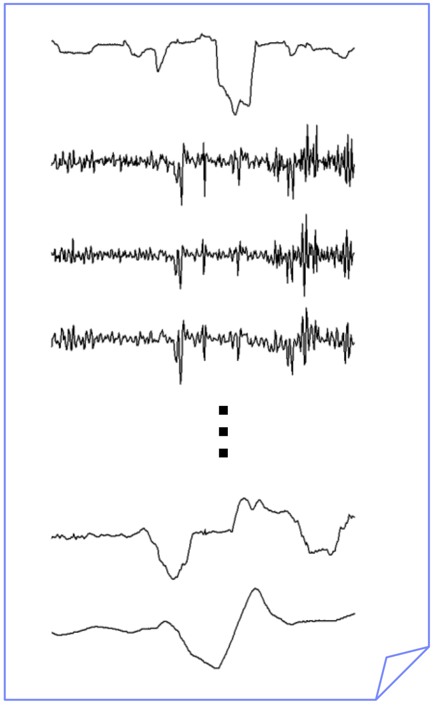
$$p(j|i) = \frac{1}{Z_i}\mathrm{e}^{-\frac{d_{ij}}{\sigma_i}} \quad \text{where} \quad Z_i \equiv 1 + \sum_{l\in\mathcal{N}_i} \mathrm{e}^{-\frac{d_{il}}{\sigma_i}}$$

This amounts to "softening" neighborhood graphs.

Experimental result and summary

E-score clearly pinpointed faulty automobile sensors, which were very hard to be detected by the human eye

reference



x_1

x_2

x_{61}

- **test data includes 3 faulty sensors**
  - due to mis-wiring error between x-, y-, and z-axes of an acceleration sensor
- **Faulty ones were clearly identified**

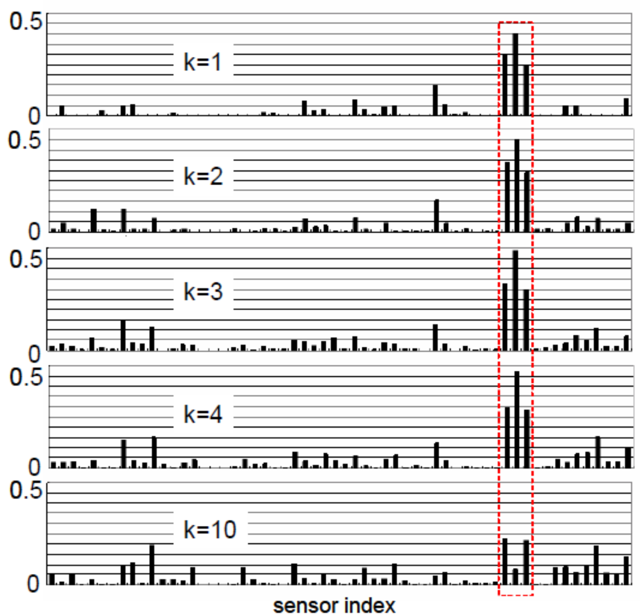
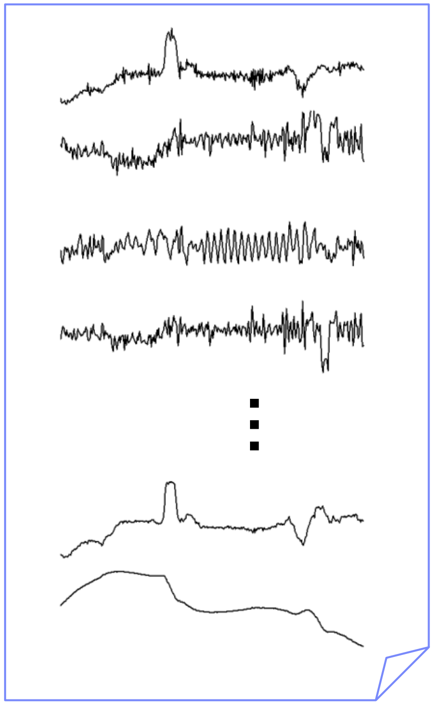


Figure 4. The k-dependence of the E-scores.

test



Summary

- **We formalize the task of change analysis**

- **We proposed the neighborhood preservation principle for change analysis**

Thanks !