グラフ時系列による
ネットワーク異常検出

数理システム
知識工学部　長沼 茂太

モチベーション

ネットワークの各ノードから得られる情報（通信記録など）の時系列データから、異常を検出したい。

- 大域的なノードの相互作用における異常
- オンライン性 → 時々刻々と平常状態が変化していくシステムに追従した異常検知
- 周期性を考慮した異常検出 → いつもの月曜日と違う
- 大規模ネットワーク

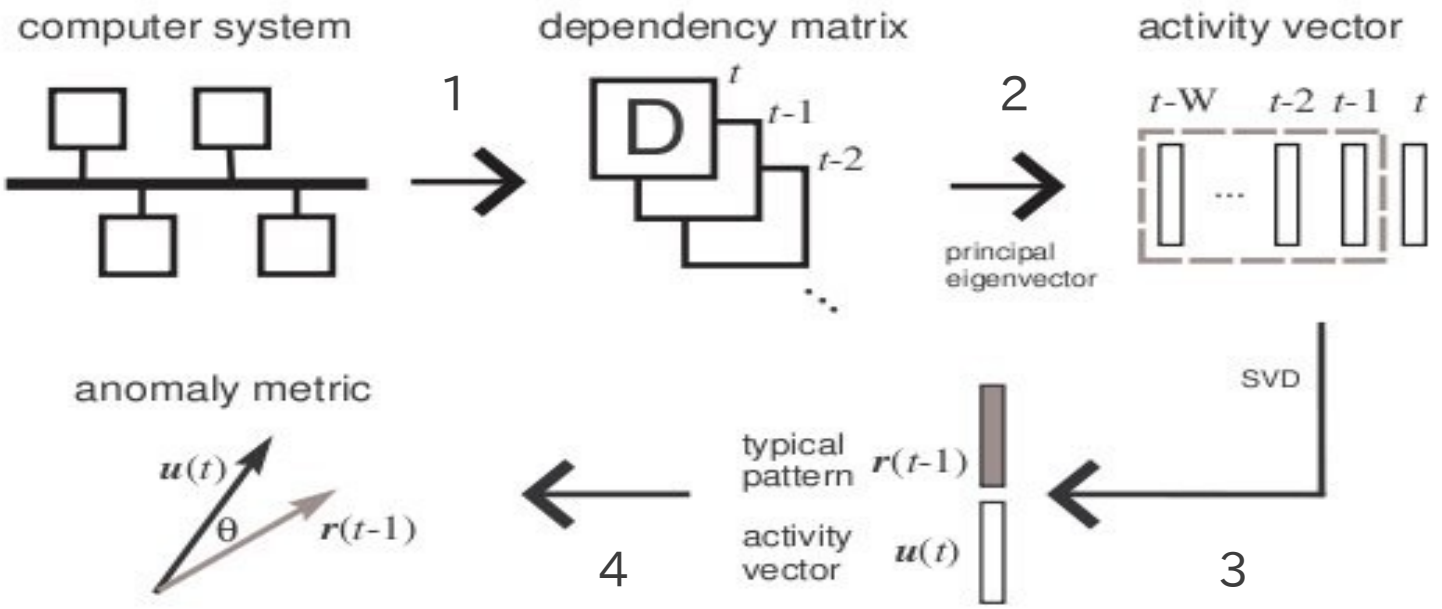
グラフマイニングによるネットワーク異常検知手法は近年盛んに研究されており、中でも以下の手法が要件に合致する手法である。この Ide & Kashima の手法を主に用いたネットワーク異常検出について紹介する。

- “Eigenspace-based Anomaly Detection in Computer Systems”
  (Ide & Kashima, 2004)

Ide & Kashima の利点

- グラフの構造解析手法
  - サイズが肥大しがちなグラフの行列を、特徴を顕在化させながら圧縮する
    - その後の解析における計算の工学的問題に対処できる。
  - 伝統的な (AR モデル + white ノイズ) のようなモデルでは対応しきれなかった非定常なデータに対しても ロバストである。
- 閾値の統計的処理
  - 従来の信号解析の手法では手入力で設定されていた閾値を、確率分布によって自動で決定していく。

概要

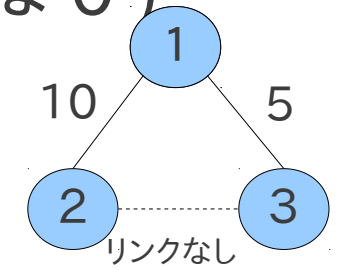


1. 各時刻の重み付きグラフ行列 (n x n の行列, n :ノード数)を得る。
2. 重み付きグラフ行列からネットワークの活動パターンを算出(u(t): n 次元ベクトル)
3. 過去のパターンから典型的パターン (r(t-1))を算出
4. u(t) と r(t-1) のずれ具合(θ)を確率分布を用いて評価

   → 通常 or 異常 判定

重み付きグラフ行列

- 本手法では、ネットワークから得られる重み付きグラフ行列 D の時系列データを解析対象とする。
  - 重み付きグラフ行列とは、サイズ N x N (N:ノード数)で、各要素は添字に対応するノード間のリンクの重みである (リンクがない場合は 0)



$$\begin{bmatrix} 0 & 10 & 5 \\ 10 & 0 & 0 \\ 5 & 0 & 0 \end{bmatrix}$$

- ネットワーク規模拡大に伴って行列のサイズは容易に大きくなってしまう。
  - 学習を効率よく行うには、必要な情報を最大限失うことなくデータサイズを圧縮した特徴量が欲しい → 活動パターン

活動パターン

- 中心性

グラフにおいて各ノードがどの程度「中心的であるか」を表現する数学的な量

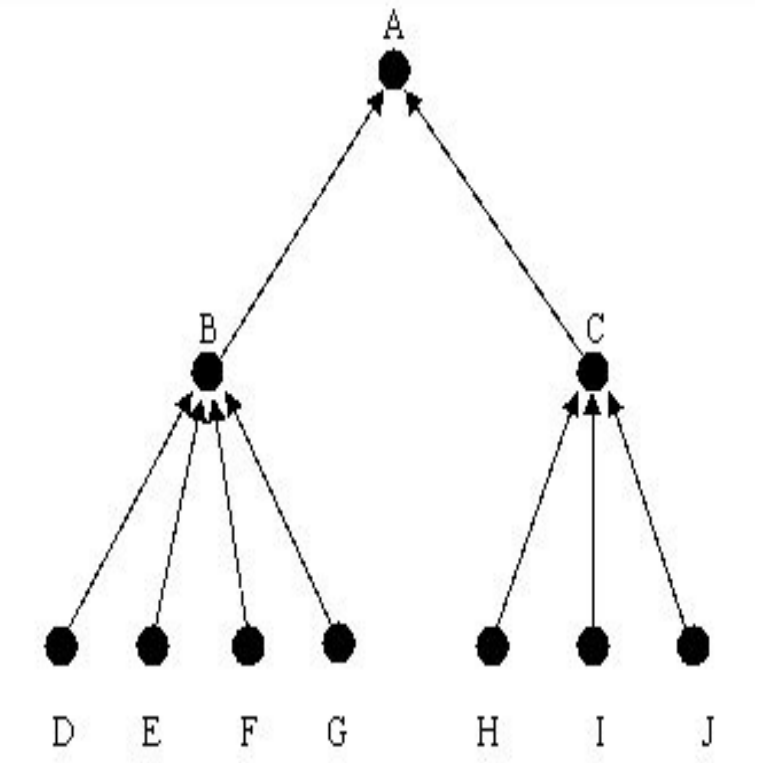
- 固有ベクトル中心性

  次数(リンク数)の高いノードとリンクをもつノードの中心性を高く評価する。

  – “信頼されている人に信頼されている人はより信頼度が高い”

  – 重み付きグラフ行列の最大固有値をもつ固有ベクトルとして算出できる。

    N x N 行列 → N 次元ベクトル



A 自体の次数は少ないが、つながるノードの次数が大きいため固有ベクトル中心性においては高い中心性を示す。

活動パターン

- 活動パターン
  - 重み付きグラフ行列としてネットワークにおける各ルーター間のトラフィック量を用い、固有ベクトル中心性を求めると、より活発に周囲のルーターとトラフィックのやりとりをするルーターの中心性が高いベクトルが求まる。

中心性ベクトル V における 1 番目の成分が他よりも大きくなる。
→ V をネットワークの活動を表すものとして活動パターンと呼ぶ。

  - 但し、V は単位ベクトル (|V| = 1)の制約から、成分値はあくまで相対的な量である。i.e. 4 が 1 よりもさらに活発になれば、V における 1 番目の成分は減少し、4 番目の成分は増加する。
    - 各ノードの活動のバランス関係を表すとも言える。

典型的パターン

- 滑走窓
  - あらかじめ定めた過去の一定範囲（窓）における活動パターンから典型的なパターンを学習する方式
  - 時間が進むにつれて、窓は新しい方向にずれていく
    - 過去のパターンは順次忘れ去られていく
- 学習法：特異値分解(SVD)
  - 代表的な機械学習手法（e.g. 潜在意味解析など）
  - 窓内の各パターンベクトルに共通の方向性を最大限強めたパターンベクトルが求まる。

    → 典型的パターンベクトル
  - 高速なアルゴリズムが存在する。(power method)

異常検出

- 異常スコア（非類似度 dissimilarity）

  現在の活動パターンと典型的パターンの成す角度 θ によって定義する。→ $z(t) := 1 - \cos\theta$ (t: 時間)

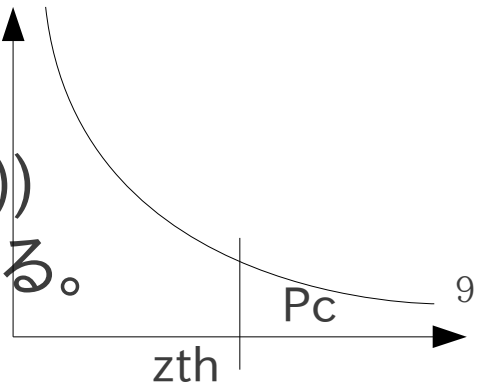
- 異常スコアの閾値

  1. vMF 分布を応用して z(t) の分布を学習
     - vMF(von Mises Fisher)分布
       - 方向ベクトルの確率分布 → パターンベクトルはまさに方向ベクトルなので最適
       - パラメータ:平均 μ 分散 σ を重み β によって過去の値を考慮しながら更新していく。

         → $\mu(t) := (1 - \beta)\,\mu(t\text{-}1) + \beta\, z(t)$

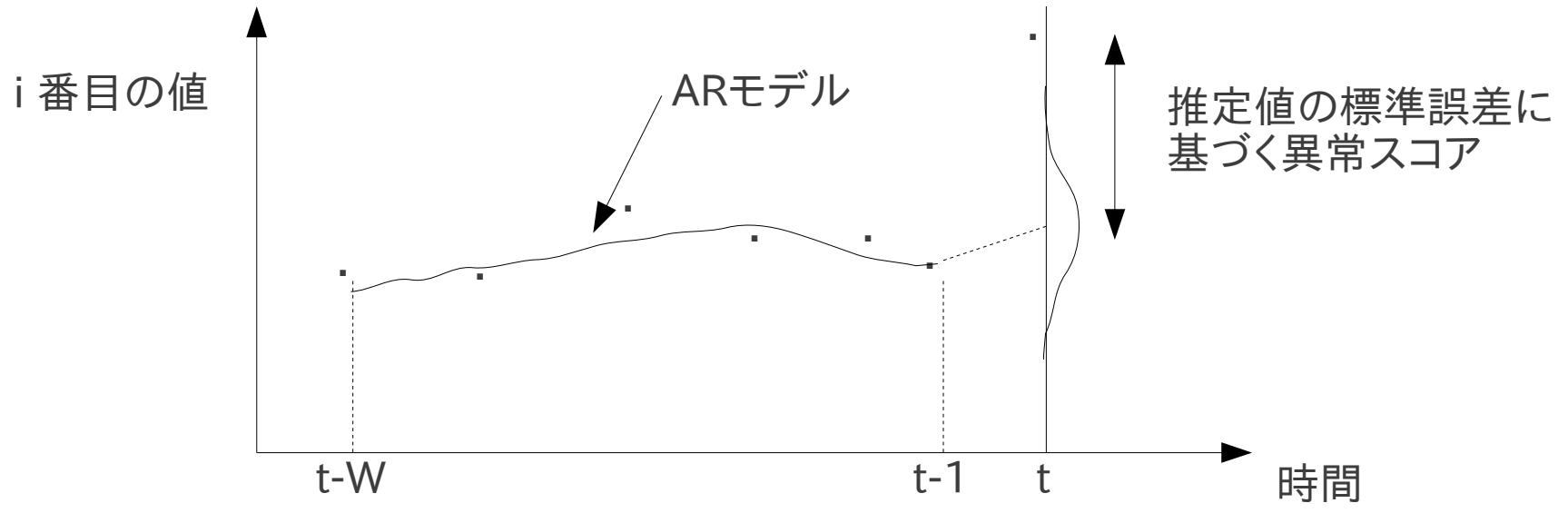
         $\sigma(t) := (1 - \beta)\,\sigma(t\text{-}1) + \beta\,(z(t) \times z(t))$
  2. 上側累積確率 Pc における値を閾値 zth とする。

     → z(t) > zth なら異常検知



異常の局所的検出

- さらに故障箇所も特定したい。
  - 活動パターンの各要素の時系列データに着目
    - 要素 i はノード i の活動の強さに対応
    - AR モデルを作成し、その推定値からのずれをもとに、ベクトルの各要素の異常スコア(i.e. 各ノードの異常スコア)を算出。



関連研究：SNN

“Computing Correlation Anomaly Scores using Stochastic Nearest Neighbors”

(T.Ide, S.Papadimitriou, M.Vlachos 2007)

- カーセンサーデータなど、各パラメータが相関をもつような多変量時系列データに対して、各パラメータごとにその相関関係を確率的近傍 (Stochastic Nearest Neighbors) で表現し、その崩れを見ることで異常検出を行う手法
  - 局所的な異常検出
  - 右図のような各パラメータの揺らぎが強いシステムでも異常解析が可能

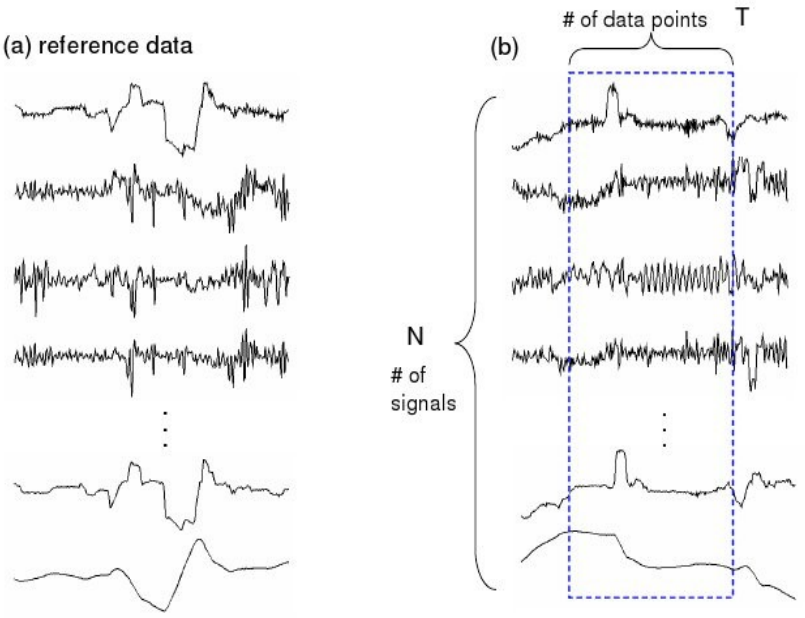
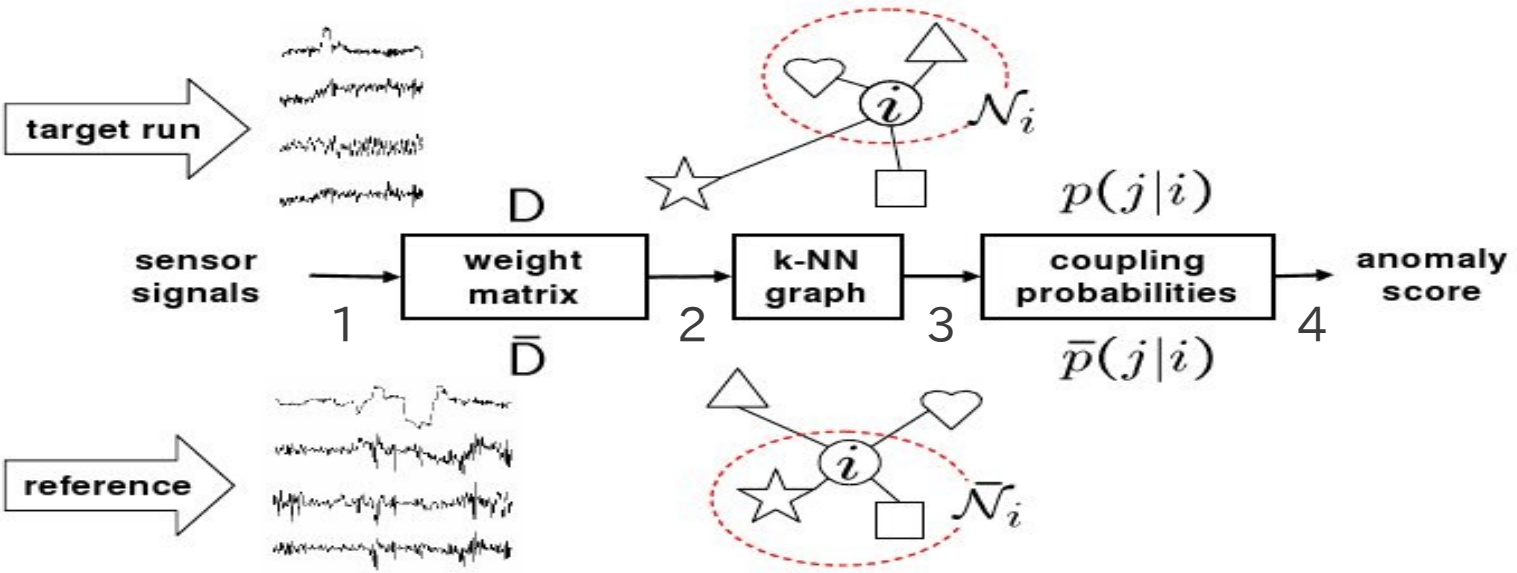


Figure 1. Problem setting. We wish to compute the anomaly score of each sensor in a target run (b) using a reference data set (a).

関連研究：SNN



- 手順
  1. ネットワーク時系列データから非類似度行列を算出
  2. 非類似度の昇順に、各パラメータの近傍となるパラメータを k 個選出
  3. 各近傍それぞれについて、ペアになる確率（リンク強度）を算出 → SNN
  4. 1→2→3 を検証データ (target) 平常データ(reference) で算出し、各パラメータの異常スコアを SNN のグラフ構造のずれから算出
- “近傍保存原理”

  システムの動的である部分、不安定なファクターのほとんどは弱い相関によるものであり、平常時であれば SNN は不変である、という著者の観測に基づく原理 → 動的なシステムでも本手法は有効。

関連研究:SNN

- 利点
  - SNNによりグラフ構造のずれが定量的に算出可能となり、個々のパラメータの異常スコアが、他のパラメータとの相関に基づいて算出できる。
  - 近傍保存原理により、試行のたびにトレンドが違うような高度に動的なシステムでも適用できる。
- 問題点
  - 近傍数 k は人による決め打ちであり、システム全体で k は一定である。→ パラメータによって最適な SNN の規模 k は異なると考える方が自然では？
  - 近傍保存原理が成り立つシステムでなければならない。→ 著者らの経験的事実であるため、どうやって体系的に確かめるのかは不明。

関連研究：EEC

“Network Anomaly Detection based on Eigen Equation Compression”

(S.Hirose, K.Yamanishi, T.Nakata, R.Fujimaki 2009)

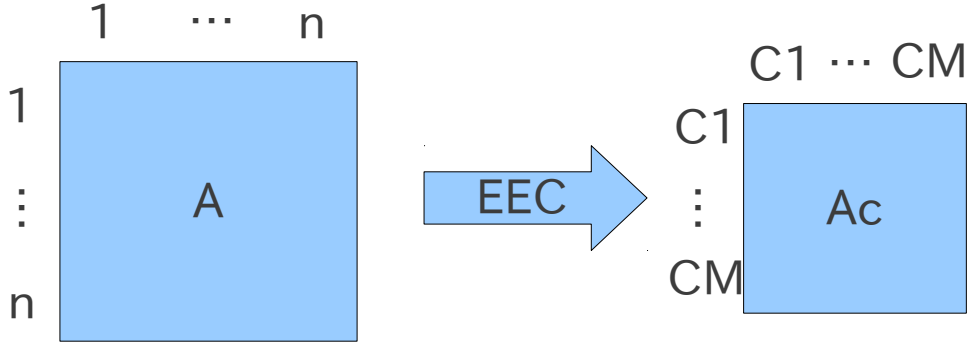
- Eigen Equation Compression という行列の圧縮手法により、大域的および局所的なネットワークの特徴を同時に定量化することが可能となり、それらの量を確率分布を用いて学習することで異常検出を行う手法
  - 大域的または局所的な異常検出手法は数多あるが、同一の枠組みで、大域的／局所的な異常を同時に検出しようとする最初の試み。

関連研究:EEC

- EECとは大まかに言えば行列圧縮の技法
  - ネットワーク時系列データの相互相関行列の各要素に対して絶対値をとった相関強度行列 A (n x n, n:ノード数)をノードのクラスタリング

    $$C = \{Ci\}\ (\ i = 1,\cdots,M\ )$$

    により、n x n 行列から M x M 行列 Ac へと圧縮する。(n >> M)
  - Aは各ノード間の相関強度を表すが、Acは各クラスタ間の相関強度を表す。



関連研究:EEC

- Acの定義
  クラスタCkへの射影行列を $P_{C,k} = \sum_{i \in C_k} e_i e_i^\dagger$
  (e_i は i 番目のみ 1, 他は 0 の n 次元ベクトル)
  Aの固有方程式を $A\psi = \lambda\psi$ とし、固有ベクトルΨのクラスタCkへの射影

$$\mathbf{C}_{C,\mathbf{k}} = \frac{\mathbf{1}}{\phi_{C,\mathbf{k}}} \mathbf{P}_{C,\mathbf{k}} \psi \ (\phi_{C,\mathbf{k}} = \sqrt{\psi^\dagger \mathbf{P}_{C,\mathbf{k}}^\dagger \mathbf{P}_{C,\mathbf{k}} \psi})$$

  ($\phi_{C,k}$ は正規化項)を用いて以下のように Acの各成分を定義する。

$$(A_C)_{k'k} = \mathbf{C}_{C,\mathbf{k}'}^\dagger \mathbf{A} \mathbf{C}_{C,\mathbf{k}} \ (\mathbf{1} \leq \mathbf{k}, \mathbf{k}' \leq \mathbf{M})$$

関連研究:EEC

- Acについて
  - M x M 次元
  - Aの固有方程式と固有値を共有する。

$$A_C \phi_C = \lambda \phi_C$$

(ϕ_C は $\phi_{C,k}$を要素とする M 次元ベクトル)

  - Acの固有値はクラスタリングに依らない。
- 各成分(k'k)はクラスタCk'に属すノードとクラスタCkに属すノードとの相関強度の重み付き和をとったものである。
  - “クラスタ間の相関強度”と考えられる。

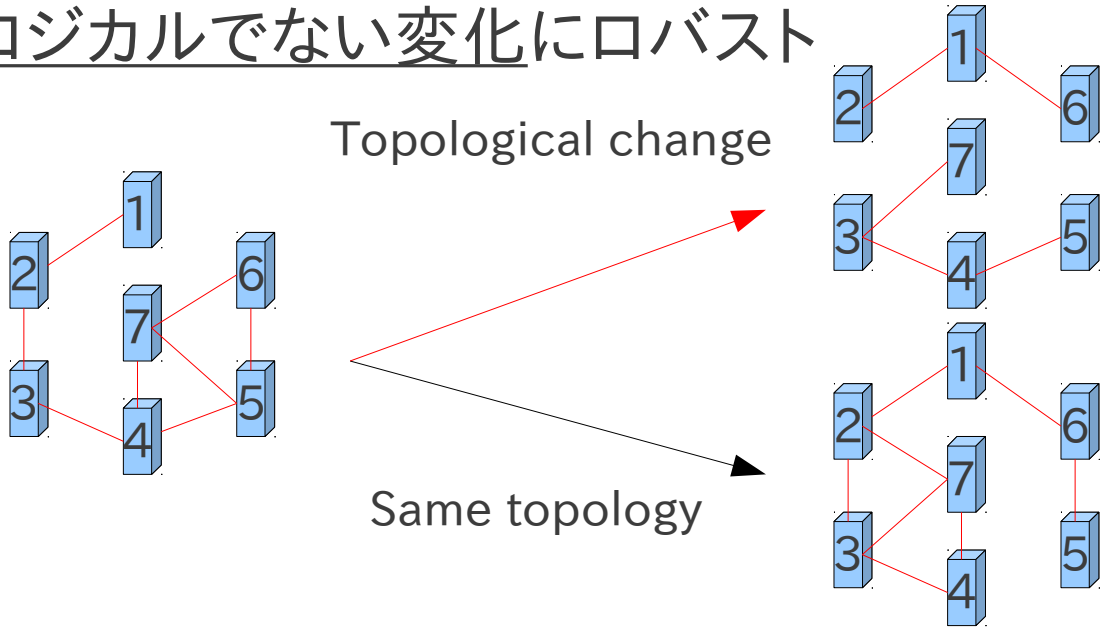
関連研究:EEC

- ノード特徴量
  - 各ノードごとに、以下のようなクラスタ数 3 のクラスタリングによる EEC を適用する。
    - C1 : 対象ノード自身
    - C2 : 対象ノードとの相関強度が ξ 以上のノードたち
      (近いクラスタ)
    - C3 : 対象ノードとの相関強度が ξ 以下のノードたち
      (遠いクラスタ)

      → Ac として 3 x 3 の行列を得る。これを各ノードの特徴量とする。
  - ノードから見たクラスタリングによるネットワークの相関構造を表す。

関連研究:EEC

- ネットワーク特徴量
  - A の固有方程式による固有値 λ を降順に上位 m 位までを並べたベクトル (m 次元ベクトル)
    - λ は各ノード特徴量 Ac で共通であり、ノードに独立な値
      → 大域的な特徴を表す量
    - λ はAの直行変換 U (ノードID並べ換え)で不変
      → トポロジカルでない変化にロバスト



関連研究:EEC

- 学習・異常スコア
  - 特徴量の分布を以下の確率分布のパラメータを推定する形で学習する。
    - ノード特徴量 → 行列正規分布: Pi
    - ネットワーク特徴量 → 多変量正規分布: P
  - 異常スコアは確率分布に基づいて、
    - ノード異常スコア Si := -log Pi
    - ネットワーク異常スコア S := -log P

    によって定義し、スコアが高い場合は対応するノードまたはネットワークが異常であると判断する。
- ノード／ネットワークの両特徴量は共に相関行列の固有方程式を元に求めているので、S かつ Si が高かった場合に、

  ”ネットワークに異常が起きており、その原因は i 番目のノードである”

  と言うことができる。

関連研究：EEC

- 利点
  - ノード特徴量とネットワーク特徴量の理論的つながりが明確なため、大域的／局所的異常を同一のフレームワークで評価できる。
    - 従来研究では独立に考えられてきたため、それぞれの異常検出結果を統合するのが困難である。
  - EECによってノード特徴量を低い次元の行列(３ｘ３)で表現できる。
  - ネットワーク特徴量はトポロジカルでない変化や局所的な変化に対してロバストである。
  - ノード数に増減があっても適用可能

関連研究：EEC

- 問題点
  - EECのためのクラスタリングが相関値に対する閾値を用いるだけの単純なもので、まだまだ工夫の余地がある。→ 今後の研究課題
  - 特徴量から異常検出に至るまでが、Ide & Kashima に比較してナイーブである。
    - 確率分布は正規分布でよいのか。
    - 確率分布のパラメータが各時間ごとに独立に推定される。

      → Ide & Kashima の手法では、過去の値の影響も考慮している。
    - 異常スコアの閾値を決定する手法について言及がない。

      → 問題毎に傾向を調査して決定しなければならない？
  - 相互相関行列を用いており、偽相関などが問題
    - ノードの関係強度を表す行列(類似度行列)であれば適用可能

      → 今後の研究課題