

Time-series clustering - A decade review

佐野 達也

このスライドは、

Time-series clustering - A decade review

(<https://www.sciencedirect.com/science/article/pii/S0306437915000733>)

を自分なりにまとめ、一部補足を加えたものです。

文章中にある [(番号)] 表記のものは、Time-series clustering - A decade review 内の References 番号を意味します。

概要

“Time-series clustering - A decade review” は、時系列クラスタリングに焦点を当て、包括的レビューを行ったものです。

この論文は、過去 10 年間 (2006~2015) の時系列手法の効率性、品質、複雑性の改善傾向に関する最新の調査を示し、将来の研究のための新たな道筋を明らかにすることを目的としています。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

- クラウドコンピューティングやビッグデータの登場により、クラスタリングなどの教師無し学習によって大量のデータから知識を抜き出す研究が活発化している。
- 時系列データのクラスタリングは多くの科学分野において、データアナリストが複雑で大規模なデータセットから貴重な情報を抽出するために使われている。
- ストレージやプロセッサの性能が向上したおかげで、実際のアプリケーションでは、長時間のデータを保持する機会ができた。(購買履歴や、株価、為替、気象データ等々)

クラスタリングとは？

- 類似のデータを関連するグループもしくは同種のグループに配置することを目的としたデータマイニング手法。
- 教師無し学習に分類される。

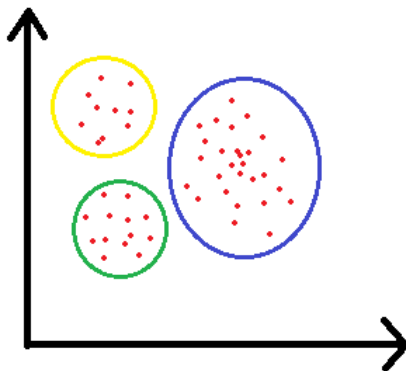


図 1: 2次元平面上でのクラスタリングの例

クラスタリングの目的

- 時系列データベースには、パターン発見によって得られる貴重な情報が含まれている。時系列クラスタリングはそのようなパターンを発見する。
- 時系列データベースはサイズが非常に大きく、人の手で識別するには限界がある。クラスタリングは類似の時系列をクラスタ化することによって構造化する。
- 探索的データ解析で最も頻繁に使われるアプローチであり、他のデータマイニング手法の前処理として使われる。
- クラスターの構造を視覚的に表すことで、ユーザに迅速にデータ構造を理解させることができる。

時系列クラスタリングは難しい

時系列クラスタリングは一般的なクラスタリングと比べ、特に以下の 2 点に苦しんでいる。

- 高次元性

時系列データの高次元性はクラスタリングの実行速度に多大な影響を与える。

- 類似度測定

時系列データ同士の次元、位相の違いや外れ値、ノイズの処理などを考慮して類似度を考えなければならない。

時系列クラスタリングの3つの種類

- whole time-series clustering(本論文ではここに焦点を当てる)
時系列集合を類似度に基づいていくつかのクラスタに分割する。
- subsequence time-series clustering
1本の(長い)時系列から部分系列を多数生成し、それらをクラスタリング。
- time point clustering
?

どのような特徴を基にクラスタリングをするのか？

- 形状ベース (shape-based)
時系列データの形を基にクラスタリングする。生の時系列データを直接扱えるため、raw-data-based なアプローチとも呼ばれている。
- 特性ベース (feature-based)
生の時系列データを低次元の特徴ベクトルへ変換する。その特徴ベクトルを用いてクラスタリングする。
- モデルベース (model-based)
各々の時系列にたいしてパラメトリックなモデルを適用しモデルパラメータを求める。そのモデルパラメータを適切なモデル距離を用いてクラスタリング。

時系列クラスタリングの4つの構成要素

- 次元削減 → chapter2
- 類似尺度・距離尺度 → chapter3
- プロトタイプの定義 → chapter4
- クラスタリングアルゴリズム → chapter5

時系列クラスタリングは、扱うデータによって、上の複数、もしくは全てを組み合わせてクラスタリングを行う。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

次元削減とは

時系列の次元削減・・・生の時系列を近似や特徴抽出を行うことにより、低次元空間へ変換すること。

時系列データは高次元であることが多いため、次元削減は頻繁に用いられる。

なぜ次元削減をするのか？ (3つの理由)

- メモリサイズを減らす

時系列データはサイズが大きく、生の時系列が全てメインメモリに収まらないときがあるので、データのメモリを減らす必要がある。

- 計算時間を減らす

時系列に用いられる距離尺度は、データの長さに大きく依存する (時系列データの距離尺度として頻繁に使われる DTW は時系列の長さに対して2乗のオーダー)。

- ノイズの影響を減らす

2つの生の時系列間の距離を測定すると、非常に直感的でない結果を得る可能性がある。

理由として、いくつかの距離尺度は時系列データの歪みに対して非常に敏感であるため、クラスタリング結果として、ノイズの類似性に基づいてクラスタリングしてしまう可能性がある。

次元削減の 4 つのタイプ

- Data adaptive 時系列をいくつかのセグメントに分け、復元誤差を最小化するしようとする方法。
主成分分析 (PCA) や、Symbolic Aggregate ApproXimation(SAX) などがある。
- Non-data adaptive データの局所特性を考慮し、それに応じて近似表現を構築する。
離散フーリエ変換 (DFT) や、離散ウェーブレット変換 (DWT) などがある。
- Model based 確率的手法でデータを削減する。
隠れマルコフモデル (HMM) や、自己回帰移動平均モデル (ARMA モデル) などがある。
- Data dictated 上の 3 つのアプローチが次元の削減率をユーザが決めることができるが、このアプローチは圧縮率は自動で決められる。

手法の比較

H. Ding et al. [91] は 8 つの異なる次元削減メソッドを 38 の時系列データセットを用いてそのパフォーマンスを調べた。
その結果、近年の次元削減メソッドの性能はほとんど差異がないことを示した。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

- クラスタリングの結果は扱う類似/距離尺度に大きく依存する。
- 時系列データには位相や長さが異なるデータなどを考慮する必要がある、時系列の類似性を適切に決定する必要がある。
- 時系列の類似度/非類似度の計算には大きく 3 つのアプローチに分かれる (次ページ)。

類似/非類似度計算の3つのアプローチ

- 時点に基づいて類似の時系列を見つける
時系列データを同じ時間点事に比較する。ユークリッド距離を用いることが多い。
- 形状に基づいて類似の時系列を見つける
データにおいて、特定のパターンが発見される時刻というのはさほど重要ではない場合がある。そのような場合、形状に基づいて類似の時系列を用いることができる。
これにより、時系列間の位相の違いを考慮することができる。
- 構造に基づいて類似の時系列を見つける
隠れマルコフモデルや ARMA 過程によって得られたパラメータの類似度に基づいてクラスタリングする。
翌日に株価が下落した後に増加傾向にある株式データのクラスタリングなど、同様の自己相関構造を持つ時系列をクラスタリングする。このアプローチは、長い時系列に適している (逆に、短い時系列には適していない)。

距離尺度に関する 3 つの議論

- 最も効果的で正確なアプローチは、動的計画法 (DP) に基づくものであることが示唆されている。しかし、DP を用いた距離尺度である DTW は時系列の長さに対して 2 乗のオーダーであり、計算量の面で問題がある。よって、いくつかの制約を課して、実行速度を緩和する研究がされているが、速度と精度の間にはトレードオフが生じる。
- 類似度に関する研究で、大きな課題の 1 つとして、変形された時系列と距離尺度との非互換性が挙げられる。例えば、時系列解析に適用される一般的なアプローチの 1 つは、周波数領域に基づいているが、これを使用して時系列間の類似性を見つけることは困難である。
- ユークリッド距離及び DTW は、時系列クラスタリングにおける類似度測定の最も一般的な尺度である。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)**
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

プロトタイプ・・・クラスタを代表するデータ。セントロイドとも呼ぶ。

プロトタイプを見つけることは、クラスタリングの中でも必要不可欠であり、k-Means 法や k-Medoids 法などの分割最適型クラスタリング (chapter5 で扱う) では、プロトタイプの質がクラスタの質に大きく影響する。

プロトタイプの定義

プロトタイプの定義には、一般的に 3 つのアプローチがある。

- データの平均をプロトタイプとして扱う
- medoid をプロトタイプとして扱う
- local search を用いてプロトタイプを定義する

以下でアプローチの内容を説明する。

データの平均をプロトタイプとして扱う

平均化・・・プロトタイプと、クラスタ内データとの距離の総和が最も短くなるようなプロトタイプを定義する。

- 長さが同じ時系列かつ、用いる距離尺度が non-elastic distance(時系列同士を同じ時刻で比較する距離尺度。ユークリッド距離など) の場合、このアプローチを用いることが多い。
- ユークリッド距離を用いる場合、同じクラスタに属するデータを単純平均したものをプロトタイプとして定義している。
すなわち、 i 番目のクラスタ C_i のプロトタイプ $V_i = \{V_{i1}, V_{i2}, \dots, V_{iT}\}$ は、以下のように定義される。

$$V_{it} = \frac{1}{n} \sum_{j=1}^{|C_i|} x_{jt}, C_i = \{x_1, x_2, \dots, x_{|C_i|}\}$$

- elastic distance においては単純平均をもちいることができず、複雑な計算が必要になる。

medoid をプロトタイプとして扱う

medoid・・・クラスタ内距離二乗和を最小にするクラスタ内のデータ (直感的には、クラスタ内で最も真ん中にあるデータをプロトタイプとしている)。

- 距離尺度が elastic distance でも non-elastic distance でもこのアプローチを適用することができる。
- 外れ値に強いという特性を持つ。

local search を用いてプロトタイプを定義する

local search . . . medoid を計算してから平均化のメソッドを用いる。

- 平均化は、ワーピングパスを基に計算する。
- 他の2つよりもプロトタイプの質が高い。

まとめ

- クラスタの精度が低くなる問題の1つは、主に分割最適型クラスタリングにおいて、プロトタイプの定義や更新方法が不十分であることである。
- 正確でないプロトタイプはクラスタリングアルゴリズムの収束にも影響し、結果としてクラスタリングの質が落ちることになる。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)**
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

アルゴリズムの6つのグループ

クラスタリングアルゴリズムは大まかに、階層型、分割最適型、モデルベース、密度ベース、グリッドベース、マルチステップ型の6つに分かれている。

以下で、1つ1つのアルゴリズムの中身、特性を見ていく。

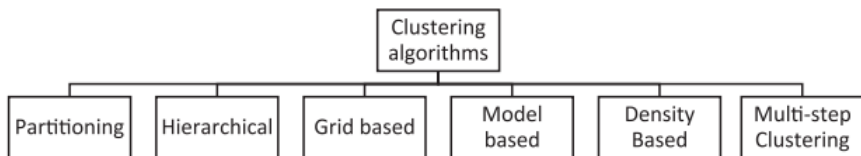


図 2: 6つのアルゴリズム 出典: "Time-series clustering - A decade review" p. 26

階層的クラスタリング

階層的クラスタリング・・・データの類似/非類似度を利用してクラスタをマージ (凝集型) もしくは分割 (分割型) をして階層を作る。

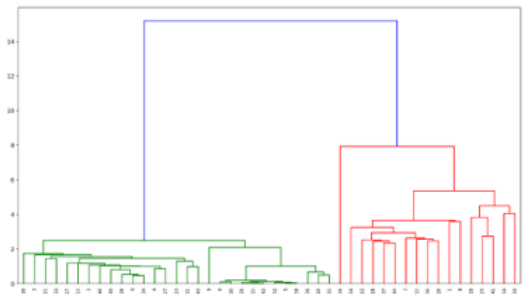


図 3: 階層的クラスタリングによって作られるデータ構造の例

階層的クラスタリング

- メリット
 - クラスタ数を事前に決める必要がない。
 - 可視化能力が高い。
- デメリット
 - 計算量がデータ数に対して 2 乗のオーダーであるため、データセットが大きい場合に適さない。

分割最適化クラスタリング

分割最適化クラスタリング・・・目的関数を決め、その目的関数を最適にする分割を求める方法。

分割最適化クラスタリングの中でもっとも有名な k-Means のアルゴリズムを以下に示す。

Algorithm 1 k-Means 法

- 1: 初期分割 $\Pi^{(0)}$ を与え、クラスタ中心 $(m_j^0)_{j=1}^k$ を求め、目的関数 $Q(\Pi^{(0)})$ を計算する。t=1 を代入。
 - 2: それぞれのデータ点 a_i について、 π_i を a_i から最も近いクラスタに更新。
 - 3: 新しい Π^t に対して、クラスタの中心点 $(m_j^t)_{j=1}^k$ を計算。
 - 4: $|Q(\Pi^{(t-1)}) - Q(\Pi^{(t)})| < tol$ を満たせば終了。そうでなければ、t に 1 を加算し、ステップ 2 に戻る。
tol はあらかじめ設定した閾値。
-

分割最適化クラスタリング

- k-Medroids 法は k-Means 法とほとんど同じであるが、プロトタイプの更新方法がデータの平均ではなく、medoid になっている (chapter4)。
- k-Means 法や k-Medroids 法はデータが 1 つのクラスタに属している (他のクラスタには属していない)。これを「ハード」、もしくは「クリスプ」なクラスタリング手法と呼ぶ。
- それに対して、Fuzzy c-Means 法や、Fuzzy c-Medroids アルゴリズムは、各クラスタに対して帰属度 (データが各クラスタに属する度合い) を持ち、「ソフト」、もしくは「ファジィ」なクラスタリング手法と呼ばれる。

モデルベースのクラスタリング

モデルベースのクラスタリング・・・クラスタに対してモデルを仮定し、それに適するようなクラスタを構成する。
統計的なアプローチを用いることが多い。

モデルベースのクラスタリング

例: 自己組織化マップ (SOM)

- 入力の高次元データを低次元 (1 または 2 次元を用いることが多い) の出力層に写像することによって入力データの類似度を出力層上で表現し、クラスタリングを行う。
- 直感的な説明は http://gaya.jp/spiking_neuron/som.htm に載っている。

モデルベースのクラスタリング

モデルベースのクラスタリングは、以下の2つのデメリットがある。

- パラメータなどをあらかじめ決めなければならない。
- 基本的に計算コストが高い。

密度ベースのクラスタリング

密度ベースのクラスタリング・・・データを密度の高い場所で分離する。

例: Density-based spatial clustering of applications with noise(DBSCAN)
各々のデータを Core 点、Reachable 点、Outlier に分け、クラスタを形成する。

アルゴリズムは次ページに乗せる。

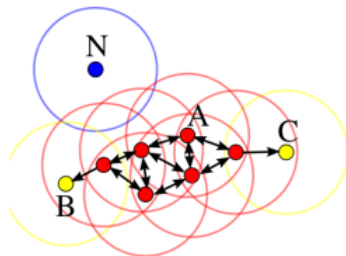


図 4: DBSCAN によるクラスタリング (赤:Core 点, 黄:Reachable, 青:Outlier) 出典:wikipedia(<https://ja.wikipedia.org/wiki/DBSCAN>)

密度ベースのクラスタリング

Algorithm 2 Density-based spatial clustering of applications with noise(DBSCAN)

- 1: 全てのデータ点を Outlier に初期化。
 - 2: すべてのデータ点に対して、半径 ϵ 以内に minPts 個の点がある場合、その点を Core 点とし、半径 ϵ 以内の Outlier の点を Reachable に変更する。
 - 3: ある Core 点から到達可能なすべての点でクラスタを形成する。Core 点から ϵ 内にある Core 点から到達可能な点も同一クラスタとする。
-

グリッドベースのクラスタリング

グリッドベースのクラスタリング・・・空間を、グリッドを形成する有限個のセルに置き換え、セル上でクラスタリングを実行する。
時系列で適用してる例がない(?)。

マルチステップクラスタリング

マルチステップクラスタリング・・・既存のクラスタリングの改善として、
以下のようないくつかのアルゴリズムが提案されている。

- 3-Phase Time series Clustering model (3PTC) [62]
- Two-step Time series Clustering(TTC) [211]

上の2つのアルゴリズムの概要を示す。

マルチステップクラスタリング

Algorithm 3 3-Phase Time series Clustering model (3PTC) [62]

- 1: 時系列に対して次元縮約 (論文中では SAX を使用) を行いクラスタリングし、サブクラスタを形成する (大雑把に計算することが目的)。
 - 2: 1 で作成したクラスタ毎にプロトタイプを計算する。
 - 3: 各クラスタのセントロイドを用いてサブクラスタをマージする。
-

マルチステップクラスタリング

Algorithm 4 Two-step Time series Clustering(TTC) [211]

- 1: 時系列集合をユークリッド距離に基づいてクラスタリング。
 - 2: ステップ1によってできたクラスタリングのサブクラスタのセントロイドを用いて、k-Medroids 法を行う (距離尺度は DTW)。
-

TTC は、ステップ2の入力データをあらかじめステップ1によって減らすことで計算時間のかかる DTW の計算回数を減らすことができる。精度の面でも従来の他のクラスタリングよりも優れていることが論文中に示されている。

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

クラスタリングの評価

Keogh や Kasetty[6] は時系列マイニングの評価は以下に従うべきだと結論づけた。

- アルゴリズムの検証は、(アルゴリズムが特定のセットに対してのみ作成されない限り) さまざまなデータセットにおいて実行できる必要がある。また、使用したデータセットは自由に利用可能である必要がある。
- アルゴリズムの実装において、注意深く行い、実装のバイアスを避けなければならない。
- 可能であれば、データとアルゴリズムは無料で提供するべきである。
- 類似性測定の新しい手法を評価する場合、ユークリッド距離などの単純で安定した尺度と比較する必要がある。

クラスタリングの評価

クラスタリング結果をどのように評価するか？

- 一般に、抽出されたクラスタの評価は、あらかじめデータラベルがない場合には容易ではなく [26]、未解決の問題である。
- クラスタの定義 (クラスタ数、クラスタサイズ、外れ値の定義など) は、ユーザ、ドメインに依存し、主観的。

しかし、人間の判断またはそのデータの生成者によってあらかじめラベル付けされたデータを用いることで、クラスタリングを評価することができる。

評価指標の種類

- クラスタリングを評価するために、視覚化、もしくはスカラー値 (こちらを中心に扱う) で評価される。
- スカラー値で評価する方法はさらに、あらかじめラベル付けされたデータとの類似度を見るが外部指標、クラスタ自体の構造の良さを評価する内部指標の2つに分類される。

評価指標の種類

以下で、外部指標、内部指標について見ていく。

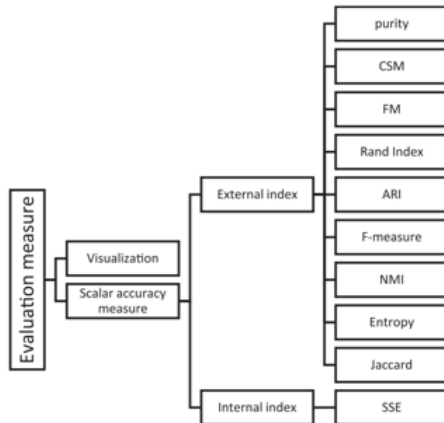


図 5: 評価指標のヒエラルキー 出典: "Time-series clustering - A decade review" p.

外部指標

外部指標・・・あらかじめ人の手によってラベル付けされたデータとクラスタリングによって付けられたラベルとの類似性を比較する。

以下で、代表的な2つの外部指標について説明する。

$G = \{G_1, G_2, \dots, G_M\}$ を人の手によってラベル付けされたクラスタ、
 $C = \{C_1, C_2, \dots, C_M\}$ をクラスタリングによってラベル付けされたクラスタとする。

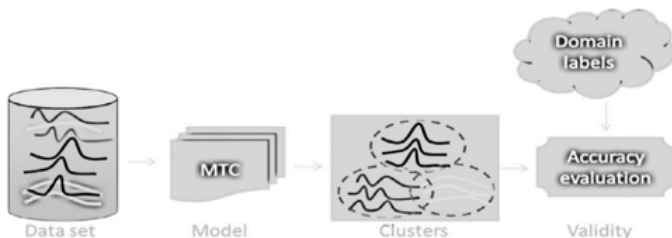


図 6: 外部指標の評価 出典: "Time-series clustering - A decade review" p. 31

外部指標

purity . . . G と C を一致率が最も高くなるように比較し、その一致率を採用する。

$$purity = \frac{1}{M} \sum_i^M \max_j |C_i \cap G_j|$$

- 値は 0~1 を取り、1 に近づくほどよいとされる。
- しかし、クラスタ数が大きいと、1 に近づきやすいという欠点がある (指標の値は、クラスタの精度だけでなく、クラスタ数にも依存してしまう)。

外部指標

Rand Index(RI)・・・任意の2つのデータが同じクラスタに属しているかどうかの確率を見る。

以下のように a, b, c, d を定義する。

- a ・・・ G で同じクラスタに属しているかつ、 C で同じクラスタに属しているデータのペア数。
- b ・・・ G で異なるクラスタに属しているかつ、 C で異なるクラスタに属しているデータのペア数。
- c ・・・ G で同じクラスタに属しているかつ、 C で異なるクラスタに属しているデータのペア数。
- d ・・・ G で異なるクラスタに属しているかつ、 C で同じクラスタに属しているデータのペア数。

外部指標

このとき、

$$RandIndex = \frac{a + b}{a + b + c + d}$$

となる。

- 値は 0~1 をとり、1 に近づくほど良いとされる。
- しかし、2 つのクラスタリング結果に相関が無いときにも値は一定にならない (0 にならない) という欠点がある。
- Adjusted Rand Index(ARI) という指標は相関が無い場合に値が 0 になるように修正されている。

内部指標

内部指標・・・クラスタリングの構造の良さを測定するために用いられる。
構造の良さの基準

- 同じクラスタに属しているデータは類似度が高い。
- 違うクラスタに属しているデータは類似度が低い。

しばしば、人の手でラベル付けすることが不可能である場合に用いられる。
以下で、代表的な2つの指標を紹介する。

内部指標

Sum of Squared Error(SSE)・・・全てのデータについて、データとデータが属しているクラスタの中心との距離の総和。

クラスタリングによって作られたクラスタを、 $C = \{C_1, C_2, \dots, C_M\}$ とすると、

$$SSE = \sum_{m=1}^M \sum_{x_i \in C_m} (\bar{x}_m - x_i)^2$$

となる (\bar{x}_m はクラスタ m のセントロイド)。

内部指標

Silhouette index(シルエット係数)・・・データ点とクラスタとの平均距離を基に計算する。

- データ点 x_i と同じクラスタ内のデータとの平均距離を a_i とする (凝集度)。
- データ点 x_i と最も近くにあるクラスタとの平均距離を b_i とする (乖離度)。

シルエット係数 s_i は以下の式で計算できる。

$$s_i = \frac{b_i - a_i}{\max\{b_i, a_i\}}$$

- 1 Chapter1 Introduction
- 2 Chapter2 Time series representation(次元削減)
- 3 Chapter3 Similarity/dissimilarity measures in time-series clustering(類似/距離尺度)
- 4 Chapter4 Time-series cluster prototypes(クラスタの代表)
- 5 Chapter5 Time-series clustering algorithms(クラスタリングアルゴリズム)
- 6 Chapter6 Time-series clustering evaluation measures(クラスタリングの評価)
- 7 Chapter7 Conclusion

近年の動向

時系列クラスタリングにおける 4 つの構成要素である次元削減、距離尺度、プロトタイプ、アルゴリズムについての近年の動向。

次元削減

さらなる次元削減メソッドは、多変量の時系列や、不均等にサンプルされたデータなどにも適用できるものがある。

距離尺度

時系列における多くの距離尺度はユークリッド距離よりも優れていない [6]。

プロトタイプ

プロトタイプに対する多くの研究は、データ平均や medoid をプロトタイプとして用いる方法より優れていない。

アルゴリズム

多くのアルゴリズムは、時間がかかるが正確性が高いか、時間はかからないが正確性が低いかのどちらかである。

時系列クラスタリングの将来

作者は以上の動向 (前ページ) を踏まえ、時系列クラスタリングにおいて、以下の結論を出している。

時系列クラスタリングの4つの側面全てにおいて改善の余地はあるかもしれない。しかし、将来の研究において、主な研究は、既存もしくは新しいクラスタリングを用いて、新たなハイブリッドアルゴリズムに取り組むことである (計算量と精度のトレードオフを考慮しつつ)。

参考

<https://www.sciencedirect.com/science/article/pii/S0306437915000733>

<https://www.sciencedirect.com/science/article/pii/S0957417413006404>

<https://www.hindawi.com/journals/tswj/2014/562194/>

http://gaya.jp/spiking_neuron/som.htm

<https://ja.wikipedia.org/wiki/DBSCAN>