

共起クラスタマイニング —数値観測量の事象系列に対する頻出パターン抽出—

稲場 大樹^{†‡} 福井 健一[‡] 沼尾 正行[‡]

大阪大学大学院情報科学研究科[†] 大阪大学産業科学研究所[‡]

1. はじめに

数値観測量の事象系列とは、波形データや位置情報、画像データのようにひとつひとつの事象が特徴量で表され、それらが系列的に発生しているデータである。そのようなデータには、頻出する興味深い共起パターンが埋もれていることが多い。共起パターンとは、系列上で互いに近接しており、かつ頻出である事象の組み合わせのことを表す。例えば、GPS データにおける停留点間の遷移パターン、気象画像における天候の遷移パターンなどが挙げられる。本研究では、上記のような数値観測量の事象系列から共起パターンを抽出する新たな手法として、「共起クラスタマイニング」を提案する。

最初に、人工データを用いて単純な共起パターン抽出法（2段階法）と提案手法との共起パターン抽出精度を比較する。次に、実データへの適用例として、燃料電池の損傷評価試験から得られた破壊音の波形データと、2011年の東日本大震災における震源リストにそれぞれ提案手法を適用し、共起パターンを抽出する。

2. 共起クラスタマイニング

共起パターン抽出の単純な方法として、2段階法[1]が挙げられる。これは、最初にクラスタリングによって、数値観測量の事象をカテゴリな事象に記号化した後、Apriori アルゴリズムに代表される記号データに対する共起パターン抽出法を適用するといった方法である。

しかしながら、クラスタリングでは時系列上においてクラスタ間の共起性が考慮されていない。そのため、実際には共起しない事象もクラスタに含まれ、反対に共起するにも関わらずクラスタに含まれない可能性がある。

Co-occurring Cluster Mining: Frequent Pattern Mining for Numerical Event Sequence

[†]Daiki Inaba, Graduate School of Information Science and Technology, Osaka University

[‡]Ken-ichi Fukui and Masayuki Numao, The Institute of Scientific and Industrial Research, Osaka University

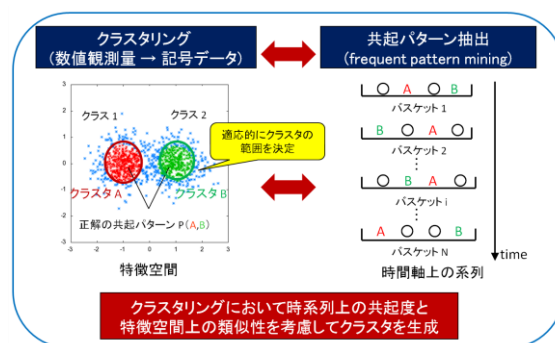


図1 共起クラスタマイニングの概念図

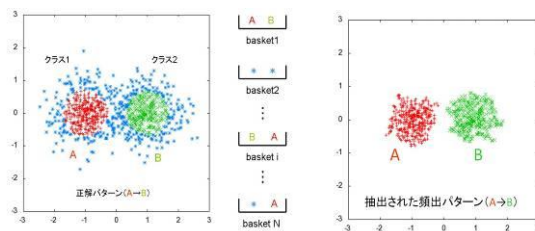


図2 人工データ（左）と提案手法により抽出された共起パターン（右）

したがって、共起パターンが適切に抽出されない可能性があるといった問題点がある。

そこで、我々が提案する共起クラスタマイニングは、時系列上におけるクラスタ「間」の共起性と、特徴空間上におけるそれぞれのクラスタ「内」の類似性を同時に考慮して、共起する2つのクラスタの範囲を適応的に決定する手法である(図1)。共起クラスタの探索には探索空間の削減のため、階層型クラスタリングによって取得されたデンドログラムを用いて探索する。

3. 人工データによる評価実験

図2に示すように、2つのクラスの分布内に共起パターンを埋め込み、2段階法と提案手法との抽出精度を評価した。2段階法では実際には共起に関係しない要素をクラスタに含んでし

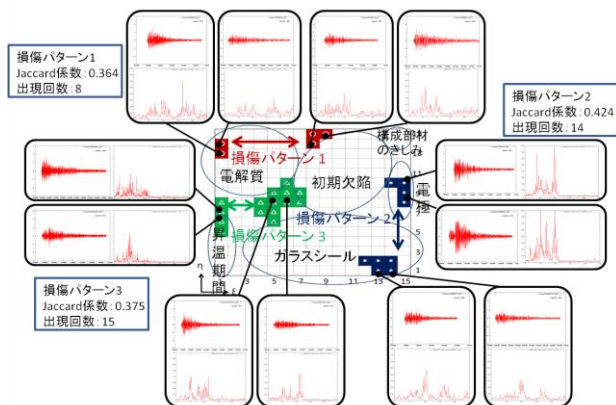


図3 抽出された燃料電池の損傷パターンの代表例。中央は自己組織化マップ(SOM) [2] による分類結果を示す。

まうが、提案手法では共起に関係した適用的にクラスタを抽出することが可能であると分かる。

4. 実データへの適用例(1)：燃料電池の損傷評価試験

本研究で用いた燃料電池は、構成部材が全て固体物質で構成されたものを用いた。燃料電池は過酷な環境下で稼働するため、物理的損傷を起し、この際に破壊音が発生する[2]。この波形は損傷の種類や部位によって特徴量が異なり、温度変化とともに系列的に得られる。この波形の系列に対して提案手法を適用した。図3はその結果である。燃料電池の専門家の評価によれば、これらの結果は妥当であるものだけでなく、構成部材間の力学的影響関係を示す興味深い結果であると評価された。

5. 実データへの適用例(2)：2011年東日本大震災における震源リスト

ひとつの地震イベントは発生時刻、震源(緯度、経度、深さ)、マグニチュード、最大震度といった数値観測量を含んでおり、今回は緯度経度を用いて地域間における地震発生の力学的な影響関係を調べる。図4はその結果の一例である。東日本大震災についての関連研究の文献[3]を参考にして評価した結果、すべり量や地殻変動量の大きさ、および強震動発生地帯であるアスペリティの分布と関係する可能性があることが分かった。

6. まとめ

本研究で我々が新たに提案した手法である共起クラスタマイニングによって、単純な手法(2段階法)と比較し、数値観測量の事象系列から



図4 抽出された地震発生パターンの一例

適切にかつ興味深い共起パターンを得ることが可能となった。本研究では、提案手法の実データへの適用例として2つ示したが、いずれも妥当そうな共起パターンが得られた。そのため、他の様々な数値観測量の事象系列に対して、本手法の汎用性が期待される。

謝辞

本研究は科学研究費補助金挑戦的萌芽研究(24650068)の支援を受けて行われました。燃料電池の損傷評価試験データは、東北大学の佐藤一永准教授および水崎純一郎先生より提供して頂きました。また、地震データの解析には気象庁の一元化震源リストを使用しました。記して感謝致します。

参考文献

- [1] Honda, R. and Konishi, O.: Temporal rule discovery for time-series satellite images and integration with RDB, *Proc. of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)*, pp.204-215 (2001)
- [2] 福井健一, 赤崎省悟, 佐藤一永, 水崎純一郎, 森山甲一, 栗原聡, 沼尾正行: 固体酸化物燃料電池における損傷過程の可視化, *日本機械学会論文集 A 編*, Vol.76, No.762, pp.223-232 (2010)
- [3] Simons, M. et al.: The 2011 Magnitude 9.0 Tohoku-Oki Earthquake: Mosaicking the Megathrust from Seconds to Centuries, *Science*, Vol.332, pp.1621-1625 (2011)