

動的タイムワーピング距離を用いた X 線天文データの類似検索

林 史尊[†] 天笠 俊之^{‡§} 北川 博之[‡] 海老沢 研[§] 中平 聡志[†]
 筑波大学情報学群情報科学類[†] 筑波大学システム情報系情報工学科[‡]
 宇宙航空研究開発機構 宇宙科学研究所 宇宙科学情報解析研究系[§]
 理化学研究所 MAXI チーム[†]

1. はじめに

近年の科学分野では膨大なデータが扱われており、その中から所望のデータを検索したり、新たな知識を発見することは極めて重要である。本研究では、天体物理学、特に X 線天体を対象とする。

X 線天体の観測データには、アウトバーストと呼ばれる、エネルギーを X 線として短期間かつ大量に放出する現象が頻繁に見られるが、その物理過程は完全には解明されていない。観測された X 線強度の変化の様子は天体やアウトバーストごとに異なるが、まれに異なる天体や同じ天体の別のアウトバーストで類似したパターンを示すものがあることが知られている。これはそれらの物理過程に何らかの共通性がある可能性を示しており、アウトバーストの起源を探る上で、大変興味深い。ただし、これまでにアウトバーストパターンの類似度を機械的に測定する手法は提案されていない。そこで本研究では、X 線天体観測データを対象に、動的タイムワーピング (Dynamic Time Warping) 法 [1] を用いて類似検索を行う手法を提案する。

ただし、アウトバーストパターンの検索では、立ち上がり(下がり)のところのみ部分的に類似したデータも検索したいという要求があり、DTW 法をそのまま適用しただけでは所望のデータを見落とす可能性がある。そこで、滑り窓 (Sliding Window) や DTW 法の改良手法である Derivative DTW 法を X 線監視装置 MAXI, BAT が観測した実データに適用し、類似パターンの発見を試みる。

2. 前提知識

2.1. 動的タイムワーピング法

動的タイムワーピング法 (Dynamic Time Warping, 以下「DTW 法」と呼ぶ) とは、二つの時系列データに対して、動的計画法に基づき、点の対応付けおよび対応付けされた 2 点間の距離の累積を求める手法である [1]。この累積距離は DTW 距離と呼ばれ、時系列データの類似度を測る尺度となる。

2.2. Derivative DTW

通常の DTW 法では一部のデータに対して望んだ点の対応が得られないという問題が知られていた。このため改良手法がいくつか提案されている。Derivative Dynamic Time Warping [2] (以下「DDTW 法」と呼ぶ) とは、そのような改良手法の一つである。時系列データに直接 DTW 法

新たな時系列データとしてあらかじめ計算しておき、それに対して DTW 法を適用するというアプローチである。

ここに、 $q = \{q_0, q_1, \dots, q_n\}$ なる時系列データがあるとする。DDTW 用に変換されるデータ Dq は次の式によって求められる。

$$Dq_i = \frac{(q_i - q_{i-1}) + (q_{i+1} - q_i)}{2}$$

ただし、 $Dq_0 = Dq_1, Dq_n = Dq_{n-1}$ とする。これにより、通常の DTW よりも値の変化に着目した点の対応付けが可能となる。

3. 通常の DTW 法を用いた場合の問題点

我々は X 線アウトバーストの類似度を算出するため、図 1 に示す天体のアウトバーストパターン間で通常の DTW 法を用いた距離算出を行った。使用するデータは、X 線監視装置 MAXI により、以下の天体からの X 線放出量を観測したものである (これらは 5 節の実験でも使用する)。図 1 にそのグラフを示している。

- GX 339-4 (橙, 黄緑, 2007, 2010 年頃のアウトバースト)
- XTE J1752-223 (赤)
- 4U 1608-52 (青)

これらは欠損値を線形補間しスムージングを施した後に自動検出したもので、更にアウトバーストデータ q のデータ値 q_i が $0 < q_i < 0$ を満たすように正規化している。

他とパターンが大きく異なる 4U 1608-52 のアウトバーストを含むペアの DTW 距離が大きくなり、パターンがよく似た GX 339-4 の二つのアウトバーストにおける DTW 距離が最も小さくなるのが期待される。他のパターンとの部分的類似を持つ XTE J1752-223 のアウトバーストを含むペアの DTW 距離は、それらの間の値に収まること

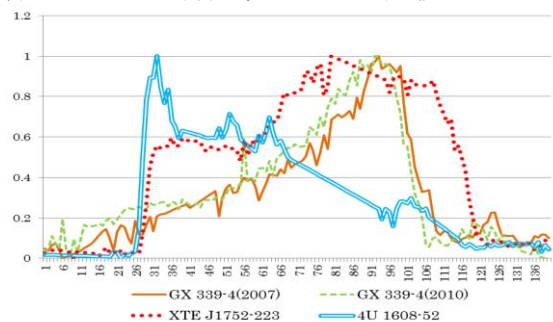


図 1: 本稿で使用する実データのアウトバースト部分

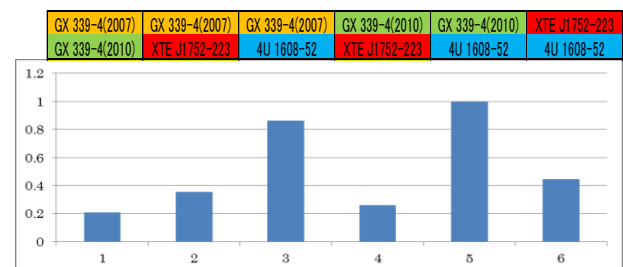


図 2: 通常の DTW 法により計算された DTW 距離

“Similarity Search of Astronomical X-ray Data using Dynamic Time Warping Distance”.
 Fumitaka[†], Hayashi[†], Toshiyuki Amagasa^{‡§}, Hiroyuki Kitagawa[‡], Ken Ebisawa[§], Satoshi Nakahir[†].
 College of Information Science, University of Tsukuba.
 Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba
 Department of Space Science Information Analysis, Japan Aerospace Exploration Agency[§].
 MAXI Team RIKEN[†].

が望ましい。

計算結果は図2のようになった。よく類似するペアほど DTW 距離は小さくなっていることが確認できる。しかし、次の点で問題があることも判明した。

我々が検出したいのは、パターン全体が類似しているパターンに加えて、アウトバーストの立ち上がり(下がり)部分のパターンが部分的に類似しているものも含む。後者の場合、通常の DTW 法では、類似しない部分が原因で DTW 距離が大きくなってしまふ。その場合、類似部分が存在するにもかかわらず、それを見逃してしまう可能性がある。また、類似している(していない)区間を部分的に見つけたいという要求にも対応することができない。このため、意図した類似パターンの検出が可能になるような手法を検討する必要がある。

4. 提案手法

4.1. 滑り窓を用いた DTW 法

3 節で述べたように、我々は部分的に類似したパターンを検出したい。そこで我々は滑り窓(Sliding Window)を利用し、部分区間に限定した DTW 法の適用を提案する。滑り窓が指定する部分区間内で DTW 距離を求め、滑り窓をずらし、次の部分区間内で DTW 距離を求める。この処理を繰り返すことによりデータのペア一つから複数の DTW 距離が得られる。そうすることで、類似・非類似部分の情報を個別に獲得できる。パターン同士の類似度は、滑り窓で得られた DTW 距離の積算値で評価する。

4.2. アウトバーストの頂点でのパターンの二分割

また、我々は、アウトバーストの頂点(最大値)でパターンを二分割した後に DTW 法を適用する手法を提案する。パターンを立ち上がり部分と立ち下がり部分に分け、それぞれについて DTW 法を適用することで、立ち上がり(下がり)の類似度が求められる。これは、DTW 法の適用範囲に制限をかけるという意味において滑り窓を用いた DTW 法の特別な場合と考えることができる。

5. 評価実験

図1の観測値データに加え、DDTW 法のために変換されたデータを用意し、これらに対し次の条件下での DTW 距離をそれぞれ計算した。

- c1. データ全体に対し、滑り窓を用いない。
- c2. データ全体に対し、滑り窓を用いる。
- c3. データを頂点(最大値)で二分する。

ただし、滑り窓は長さ 150 のデータに対し基本窓幅 50, 基本ずらし幅 10 を設定し、実際のデータ長に合わせて伸縮させている。すなわち、各々のペアに対して 11 個の DTW 距離が得られる。

実験結果の一部を図3~5示す。ただし、条件 c2, c3 の結果図4, 5については、各部分区間での DTW 距離の積算値を示している。注目すべき点は図6のグラフで、これはペア6(図2の赤と青のグラフ)に対して条件 c2 で求めた、滑り窓をずらしていった時の DTW 距離の遷移を表す。図4では積算 DTW 距離が大きくなっているが、図6を見れば、立ち上がり部分の類似を知ることができる。

6. まとめと今後の課題

本稿では、X 線アウトバーストの類似度を DTW 法を用いて計算する手法を提案し、それらの比較検討を行った。

滑り窓や部分区間に対して DTW 法を適用することによって、X 線アウトバーストのパターンの特徴を考慮に入れた比較が可能となることを示した。

今後は、提案手法を元に、知られている X 線天体のアウトバーストパターンの類似度を網羅的に探索するとともに、詳細な分析を行い、大量データから類似したアウトバーストパターンの自動検出を可能にする手法を開発する予定である。

参考文献

[1] Sakoe H, Chiba S(1971).
 "A Dynamic Programming Approach to Continuous Speech Recognition."
 In Proceedings of the Seventh International Congress on Acoustics, volume 3, pp.65-69.
 Akad[em]iai Kiadó, Budapest.

[2] Eamonn J. Keogh, Michael J. Pazzani (2001).
 "Derivative Dynamic Time Warping".
 SDM'2001 Accepted Papers.

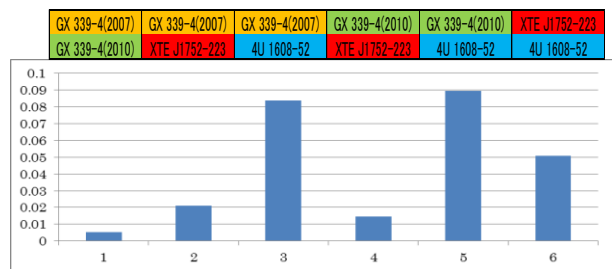


図3: 条件 c1 における, DTW 用変換データでの DTW 距離 (横軸はデータのペア, 縦軸は DTW 距離)

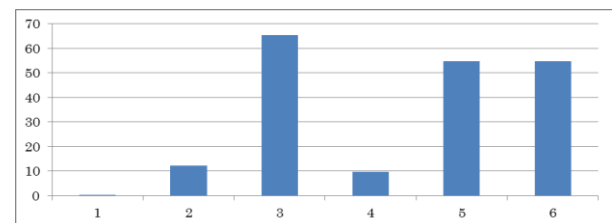


図4: 条件 c2 における, 観測値データでの積算 DTW 距離

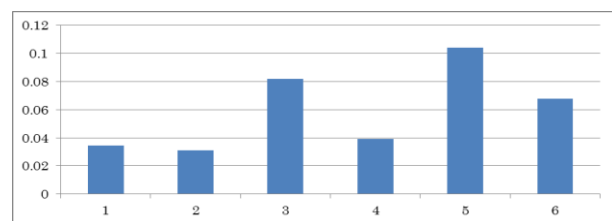


図5: 条件 c3 における, 観測値データでの合計 DTW 距離

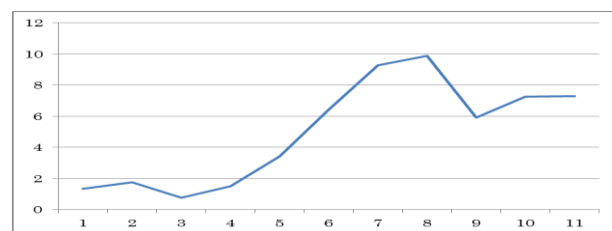


図6: 条件 c2 下, ペア6における観測値データでの DTW 距離の推移 (縦軸は滑り窓をずらした回数)