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 [*1]　天笠俊之 [*2,*3]　北川博之 [*2]　海老沢研 [*3]　中平聡志 [*3]

Similarity Search of Astronomical X-ray Data using DynamicTimeWarping Distance

Fumitaka HAYASHI[*1], Toshiyuki AMAGASA[*2,*3], Hiroyuki KITAGAWA[*2], Ken EBISAWA[*3] and Satoshi NAKAHIRA[*3]

Abstract

Explosive growth of data volume in scientific domains has caused demand of machine processing of the massive scientific data. In this paper we propose several schemes of similarity search over astronomical X-ray data for X-ray outbursts. Specifically, we first detect outbursts from the original data, followed by smoothing for reducing noise and normalization. Having detected outburst patterns, we apply dynamic time warping (DTW), which is knowntoberobustagainsttimescalingtoevaluate similarities betweentwooutburst patterns.We proposeseveral variationsbasedonDTWbytaking featuresoftheoutburst patternsand requirementsforthe similaritysearch;we apply derivative DTW, which is a variant of DTW, and apply sliding windows to evaluate partial similarities. We evaluate feasibility of the proposed schemes by experiments using real X-ray astronomy data.

Keywords: Astronomical X-ray Data, Similarity Search, DynamicTimeWarping.

概　要

科学分野で扱われるデータ量は爆発的に増加しており，膨大なデータに対する機械的処理への要求は極めて高い．論文では，天体物理学における X 線天体の X 線アウトバースト現象を対象に，その観測データの類似検索を行う手法を提案する．具体的には，観測データからアウトバースト部分の自動検出を行なう．得られたデータに対し，ノイズ除去を目的とした平滑化と正規化を施した上で，動的タイムワーピング（DTW; DynamicTimeWarping）法を適用する．DTW 法は，長さが異なる時系列データに対しても適用可能であるだけでなく，時間軸方向のスケーリングに対しても頑健なマッチングを行うことが距離である．さらに，X 線アウトバーストの持つ特徴や，類似検索に対する要求を考慮し，DTW 法の改良手法である Derivative DTW 法や，DTW 法に滑り窓を適用した手法など，いくつかの新たな手法を提案する．さらに，実データとの比較によってその有効性を評価する．

キーワード：X 線天文データ，類似検索，動的タイムワーピング．

*1 筑波大学大学院システム情報工学研究科（Graduate School of Systems and Information Engineering, University of Tsukuba）
*2 筑波大学システム情報系（Faculty of Engineering, Information and Systems, University of Tsukuba）
*3 宇宙航空研究開発機構宇宙科学研究所（Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency）

1 はじめに

近年，科学分野で扱われる観測データやシミュレーションデータは膨大なものとなっている．そのため，膨大なデータに対して高速な検索や分析の手段を提供することは，科学分野の進展のために不可欠なものとなっている．天文分野においても，日々蓄積される観測データへの対応は重要な課題となっている．本研究ではX線天体の観測データを取り上げる．X線天体とは強力なX線を放出する天体であり，その例にはブラックホール，中性子星がある．これらの天体が放出するX線の強度を観測すると，短期間に強度が大きく上昇する現象が観測される．この現象は**X線アウトバースト**と呼ばれ，このとき，天体は重力エネルギーをX線として大量に放出している．

X線アウトバースト現象の物理過程は，完全には解明されていない．例えば，観測されるX線強度の変化の様子は，天体や観測されるアウトバースト毎に異なるが，まれに類似した波形を示すことが明らかにされている[1)]．これは，アウトバースト現象の物理過程に何らかの共通性があるという可能性を示唆しており，興味深い．このため，波形の類似するアウトバーストを検出することは，アウトバースト現象の起源を解き明かす上で重要である．例えば，図1は，全天X線観測装置MAXI[2)1]によって観測された異なる天体のアウトバーストである．X線強度に関する正規化を行うと，X線強度がなだらかに上昇した後，急激に落ちるところなど，共通点があることが分かる．

X線天体の観測を行っているセンサーはMAXIの他にもあり，数多くの観測データが蓄積されつつある．このため観測されるデータは膨大であり，類似する波形を人手で発見するのは極めて困難である．

一方，時系列データのに対する検索やマイニングには

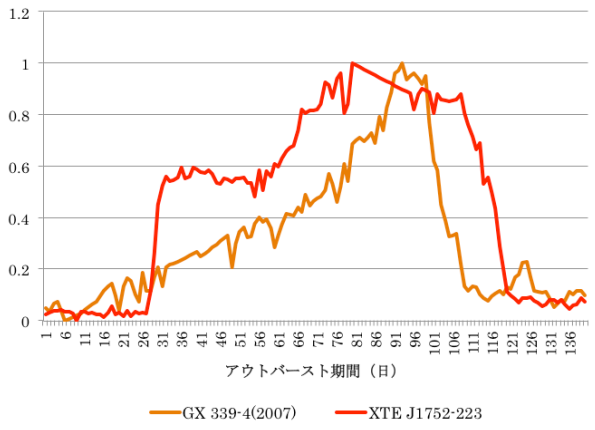


図1 波形の類似したX線アウトバースト．

1 http://maxi.riken.jp/top/

多数の研究が存在する．特に，類似検索については，フーリエ変換[3)]，ウェーブレット変換[4)]，主成分分析[5)]など多数の方法が提案されており，分子生物学，経済学，音声認識，移動体分析などに広く応用されている．時系列データの類似度を検討する際には，対象となるデータの特徴や検索に対する要求などを考慮し，適切な手法を選択することが重要である．

本研究では，X線アウトバーストの時系列観測データを対象に，類似検索を行なう手法を提案する．具体的には，オリジナルの観測データに対し，アウトバースト部分の自動検出を行なう．次に，抽出された時系列データに対して，ノイズ除去を目的とした平滑化と正規化を施す．得られた時系列データに対して，時系列データに対する距離関数を適用し，両者の類似度を評価する．本研究では，動的タイムワーピング（DynamicTimeWarping）法（以下，DTW法）[6)]を適用する．これは，動的計画法に基づく時系列データ上の距離関数であり，上で述べた関連手法に比べて，長さの異なる時系列データに対しても適用可能である，時間軸方向のスケーリングに対して頑健もマッチングを行うことが可能である等の特徴を持つ．X線アウトバーストの時系列観測データは，長さが一定でなく，時間軸方向のスケールも一定していないという特徴があるため，DTW法を用いることとした．さらに，DTW法の改良手法であるDerivativeDTW（DDTW）法[7)]の適用も検討する．

一方，アウトバーストの波形の類似検索には，例えば立ち上がり（下がり）の勾配が類似しているなど，波形のうち部分的な類似性に着目した検索に対する要求が存在すると考えられる．これに対応するため，DTW法（DDTW法）に滑り窓を適用することで，時系列データのうち部分的な類似性を考慮した検索を可能にすることを提案する．また，時系列データをX線強度が最大になる点の前後で二分割し，それぞれのについてDTW法（DDTW法）で類似度を評価する手法についても検討する．最後に，以上の手法を実データに適用し，その有効性を評価する．

本論文の構成は以下の通りである．2節では，データの前処理とアウトバースト部分の検出方法を説明する．3節で提案手法の詳細を述べ，4節で提案手法を実験により評価する．5節はまとめである．

2 アウトバーストの自動検出と前処理

X線観測データにおいて，アウトバーストが検出されている部分は全体のうちごく一部である．また，一つの天体から複数のアウトバーストが検出されるため，アウ

トバーストが検出されている部分だけを自動的に抽出する処理が必要である．また，生データにはノイズが乗っており，またX線強度も天体によって異なるため，ノイズ除去や正規化など適切な前処理を施す必要がある．以下では，アウトバースト部分の自動検出と前処理について説明する．

2.1　アウトバーストの自動検出

ある観測データ $x = \langle x_0, x_1, \ldots, x_{n-1} \rangle$ におけるアウトバーストの自動検出は，次の手順で行う．

1. **欠損値の補完** さまざまな理由により，観測値が欠損している場合がある．その場合は，前後の観測値に基づく線形補完を行い，欠損値を補う．
2. **基準X線強度の設定** 観測値であるX線強度の最大値 $x_{max} = max(x_0, x_1, \ldots, x_{n-1})$ と最小値 $xmin = min(x_0, x_1, \ldots, x_{n-1})$ を求める．次に，x_{min} と x_{max} の間を k 等分（本論文では $k = 10$）し，各区間における観測値出現頻度を計算する．最も出現頻度の大きい区間について，その区間に該当する観測値の平均を**基準X線強度** x_{base} とする．
3. ***m* 点平均の計算** m 点毎の観測値の平均 $y_j = \frac{\sum_{i=mj}^{m(j+1)-1} x_i}{m} \ (j = 0, m, 2m, \ldots)$ を算出する．本研究では $m = 20$ とした．
4. **アウトバーストの検出** m 点平均が二つ以上連続して $x_{base} + \frac{x_{max} - x_{base}}{k} \leq y_j$ となる区間をアウトバーストとして検出し，y_j に対応する観測データ x_i を抽出する．なお，抽出する観測値には，上記区間に加えて，前後の m 点（合計 $2m$ 点）も含めることとする．すなわち，$y_j, y_{j+1}, \ldots, y_{j+l}$ が連続して上記の条件を満たしていたとすると，y_{j-1} と y_{j+l+1} に対応する観測値も抽出する．

2.2　平滑化

観測値にはノイズが含まれているため，平滑化によってその影響を低減する．第2.1節では，観測値の強度の推移からアウトバーストを検出することを目的としていたため単純な m 点平均を計算していたが，ここでは株価のチャート分析などにも利用される n 点線形加重移動平均を用いる．***n* 点線形加重移動平均**は，以下のように計算される．データ列 $q = \langle q_0, q_1, \ldots, q_{n-1} \rangle$ が与えられたとき，平滑カゴのデータ列 $Sq = \langle Sq_0, Sq_2, \ldots, Sq_{n-1} \rangle$ は図2に示した式に従って求められる．

2.3　X線強度の正規化

観測されるX線の強度は天体毎に大きく異なるため，最小値が0，最大値が1となるよう，[0;1] 区間に正規化する．図3は，GX339-4，XTEJ1752-223，4U1608-52の三つの天体のデータから提案手法で抽出したアウトバーストのデータに平滑化を施したデータをプロットしたものである．なお，GX339-4については，2007年のデータと2010年のデータがあり，それぞれGX339-4(2007)とGX339-4(2010)としている．正規化前（上）では，4U1608-52のX線強度が他に比べて強く類似性が判別しづらいが，正規化によって，パターンの特徴による類似性がはっきりすることが分かる（下）

3　アウトバーストの類似検索

本節では，アウトバーストの類似検索手法を述べる．前処理を施した二つのアウトバーストの時系列データに対して類似検索を行なう場合，一般的には両者の（非）類似度を距離によって評価する．距離が小さいなら，両者はより類似していることになる．このとき，最も適切な距離尺度を選択することが重要である．例えば，最も馴染み深いユークリッド距離は，比較対象となる時系列の次元数（要素数）が等しくなければ適用することができない．このため，長さの異なる時系列データの間の距離尺度が多数提案されている．本研究では，その中でも広く用いられているものの一つである動的タイムワーピング（DynamicTimeWarping）法（以下，DTW法）[6]，およびその改良手法であるDerivativeDTW（DDDTW）法[7]を採用する．

$$
Sq_i = \begin{cases} q_0 & (i = 0) \\ \frac{1}{weightsum} \left\{ \frac{n+1}{2} q_i + \sum_{k=1}^{N} \left(\frac{n+1}{2} - k \right) (q_{i-k} + q_{i+k}) \right\} & (0 < i < m) \\ q_m & (\mathrm{i = m}) \end{cases}
$$

$$
N = \begin{cases} i & (i < (n+1)/2 - 1) \\ n - i & (i > m + 1 - (n+1)/2) \\ \frac{n+1}{2} - 1 & \text{otherwise} \end{cases}
$$

$weightsum$: 重みの合計.

図2 *n* **点線形加重移動平均.**

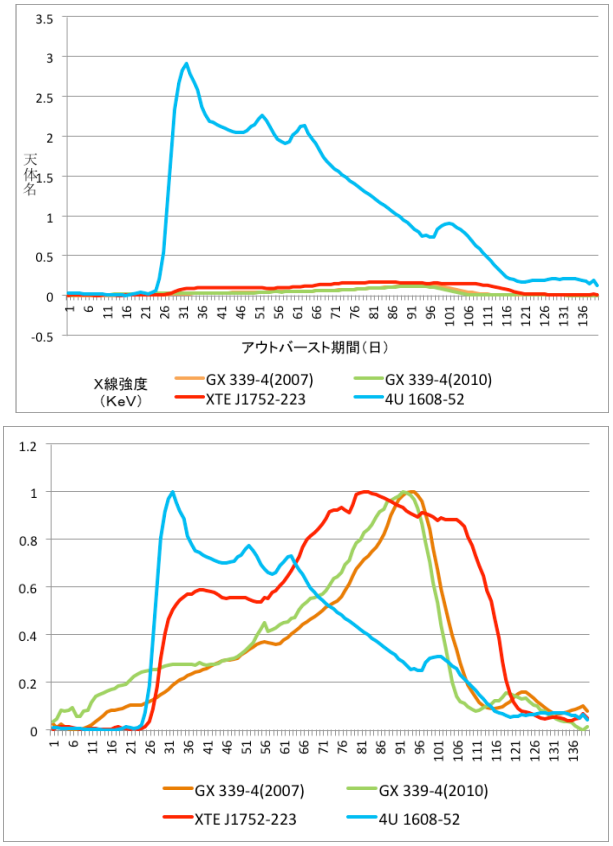


図 3 正規化前（上）／正規化後（下）.

以下では，まずその概要を説明する．続いて，時系列データの部分的な類似性に着目した類似検索を可能にするため，滑り窓を用いる方法と，時系列データを X 線強度が最大となる点の前後で二分割する方法を説明する．

3.1 基本事項

動的タイムワーピング（DTW）法

動的タイムワーピング（DynamicTimeWarping;DTP）法 [6] とは，二つの時系列データの最適なマッチングを動的計画法 [8] によって求め，それのマッチングに基づいて距離を計算する手法である．その特徴として，長さの異なるデータの比較にも利用でき，また時間軸方向のスケーリングに対しても頑健という性質を持っている．このため，音声認識など幅広い分野で用いられている [9,10]．

二つの時系列データ：

$$x = \langle x_0, x_1, \ldots, x_i, \ldots, x_{n-1} \rangle$$
$$y = \langle y_0, y_1, \ldots, y_j, \ldots, y_{m-1} \rangle$$

に対して，DTW 距離 $D(x, y)$ は図 4 に示した漸化式によって求められる．ここで，$d(x_i, y_j)$ は x_i と y_j の距離であり，差の絶対値に対して単調増加性があれば任意の関数を利用して良い [11]．本研究では，$d(x_i, y_j) = |x_i - y_j|^2$ とした．また，DTW 距離の計算は動的計画法によって行なうことができる．

図 5 に，図 3（下）で示した四つのアウトバースト（GX339-4(2007)，GX339-4(2010)，XTEJ1752-223，4U1608-52）に DTW 法を適用して距離を計算した結果を示す．元の波形を見ると，GX339-4(2007) と GX339-4(2010) が最も類似している．次に，この二つと XTEJ1752-223 が，最大 X 線強度までなだらかに上昇し，その後急激に落ちているという点で類似性が見られる．DTW 法による距離もそれを反映していることが分かる．

DerivativeDTW 法

既に述べたように，DTW 法は時系列データの距離を計算する際，極めて有用であるが，いくつか欠点も指摘されており，それに対応するための改良手法が提案されている [7,9,12-14]．その中でも最も大きな問題の一つが，スパイクのような急激な変化があったときに，その点の周辺で二つの時系列間の点同士の対応関係がいびつになってしまい，正しく距離が評価されないという問題である．この問題に対応するため，Keogh らは DerivativeDTW 法を提案した [7]．DerivativeDTW 法の基本的なアイデアは，各点の距離ではなく，変化量を比較しようとするものである．このため，時系列データに対して以下の前処理を行なう．変換対象のデータを $q = \langle q_0, q_1, \ldots, q_{n-1} \rangle$ とすると，以下の式によって得られる Dqi が変換後のデータとなる．

$$D_{q_i} = \begin{cases} \frac{(q_i - q_{i-1}) + (q_{i+1} - q_{i-1})/2}{2} & (0 < i < n) \\ D_{q_1} & (i = 0) \\ D_{q_{n-1}} & (i = n) \end{cases}$$

$$D(x,y) = \gamma(m,n)$$
$$\gamma(i,j) = \begin{cases} d(x_0,y_0) & (i=j=0) \\ d(x_0,y_j)+\gamma(0,j-1) & (i=0, j>0) \\ d(x_i,y_0)+\gamma(i-1,0) & (i>0, j=0) \\ d(x_i,y_j)+min\{\gamma(i-1,j-1),\gamma(i-1,j),\gamma(i,j-1)\} & (i>0, j>0) \end{cases}$$

図 4 DTW 距離 .

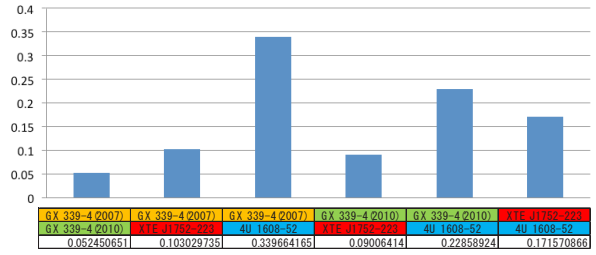


図 5　DTW 法による距離計算の例.

注目する点自身を含む周囲 3 点における平均変化量が，新たな時系列データとして生成される．例として，図 3（下）に示した GX339-4(2007) に対して上記の変換を施した例を図 6 に示す．変換されたデータ同士に対して，通常の DTW 法を適用することで，DDTW 法による距離計算ができる．図 7 に，図 3（下）で示した四つのアウトバースト（GX339-4(2007)，GX339-4(2010)，XTEJ1752-223，4U1608-52）に DTW 法を適用して距離を計算した結果を示す．DTW 法の結果に比べて，4U1608-52 との距離が相対的に上がっていることがわかる．これは，通常の DTW 法では，4U1608-52 の最大 X 線強度の前後において，他の点とのいびつな対応付けが見られるのに対し，DDTW 法ではそれが解消されるためである．

3.2　アウトバーストの部分的な類似性に着目した類似検索

DTW 法あるいは DDTW 法をそのまま用いることによって，アウトバーストの波形全体による類似度の計算

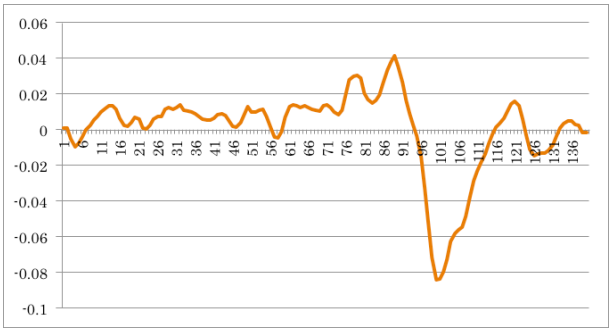


図 6　DDTW 法によるデータ変換の例.

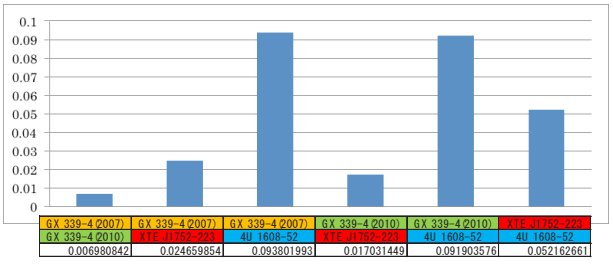


図 7　DDTW 法による距離計算の例.

が可能である．しかしながら，時系列データ全体ではなく，部分的な類似度を考慮した検索を行ないたい場合も考えられる．このため以下では，滑り窓（slidingwindow）の適用と，時系列データを X 線強度が最大になる点で前後に二分割して，前半（後半）の類似度を評価する方法を提案する．これにより，波形全体のうち 50% 以上が類似しているアウトバースト，X 線強度の上昇（下降）のしかたが似ているアウトバーストといった検索が可能となる．

滑り窓の利用

滑り窓（slidingwindow）とは，窓（window）と呼ばれる固定幅の部分系列に関して距離計算を行ない，その後，窓をずらし幅 s だけずらして，同様の計算を繰り返し行なう計算手法のことである．これに DTW（DDTW）法を適用する．その結果，部分系列（窓）間の距離が，窓をずらした回数だけ得られる．これは，部分系列間の距離が波形の先頭から末尾に向かってどのように推移するかを示している．波形全体の距離は，得られた窓内の距離の総和で評価できる．また，ある閾値以下の距離に対応する窓の位置から，波形が部分的に類似している区間を見つけることも可能である．通常の滑り窓法は，窓幅 w とずらし幅 s は比較対象の二つの時系列データに対して共通である．しかし，時系列データ全体の長さが極端に異なる場合，考慮が必要である．本研究では，w と s を，比較対象の波形の長さに対する相対値で与えることでこれに対応した．

最大点の前後による二分割

アウトバーストの X 線強度の変化は，基本的に最大強度まで上昇し，その後減少することから，最大値の前後によって二分割することが可能である．このとき，最大強度まで（以降）の X 線強度の上昇（下降）のしかたに着目した類似度の評価を行ないたい場合などが考えられる．これには，波形を最大値の前後で二分割し，それぞれについて DTW（DDTW）法を適用すれば良い．

4　評価実験

提案手法の有効性を検証するため，実データによる実験を行なった．

4.1　実験環境およびデータセット

実験は，CPU に 4 コア Intel(R) Xeon(R) CPU E5310 (1.60GHz)，メモリ容量 5GB の PC を使用した．言語は，DTW 法以外の実装と計算には JVMver-sion1.6.030 上で

動作する Scalaver.2.9.1 を利用した．DTW 法の計算には，FastDTW[15,16] に含まれる DtwTest.java を使用した．データセットとしては，人工衛星 Swift に搭載された硬 X 線モニター BAT によって観測された公開データを使用した [17]．BAT の観測対象となっている X 線天体数は 951 あるが，その中で X 線強度が強い 155 の天体を対象とした．これらのデータに対して，2 節で説明したアウトバーストの検出を行ない，平滑化および正規化を行なった．なお，一つの天体から複数のアウトバーストが検出された場合は，その中から期間の長い二つを選択した．結果として 153 件のアウトバーストを検出した．図 8 に，各天体について何件のアウトバーストが検出されたかをまとめた図を示す．抽出されたアウトバースト全ての組合せ 11,268 通り（=153*152/2）について，以下に説明する手法で距離を計算し，その上位について結果を評価した．

4.2 実験結果

DTW 法，DDTW 法による類似検索

DTW 法および DDTW 法によって，類似したアウトバーストが検索できるかどうかを評価した．距離としては DTW 法，DDTW 法を用い，上位のデータを比較した．表 1，表 2 に，それぞれ DTW 法，DDTW 法による類似検索結果の上位 20 件を示す．表において，1，3 列目はアウトバーストが観測された天体名，2，4 列目は当該天体の中で観測された何番目のアウトバーストであるか，5 列目は類似度（DTW 距離）を表している．なお，ここでの類似度は，値が小さければ小さいほど二つの時系列データが類似していることを示している．

また，上位 6 件の実際の時系列データのプロットを図 9 に示す．なお，左上から右の順に，1 位，2 位の順に並べてある．これから分かるように，どちらの手法とも，類似した時系列データが検索できていることが分かる，DTW 法の 2 位には，類似していない時系列データ

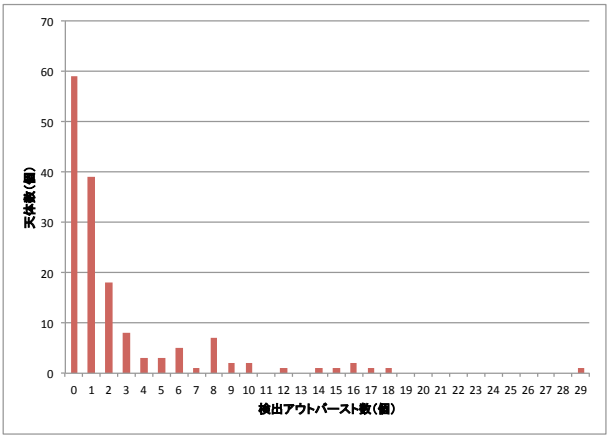


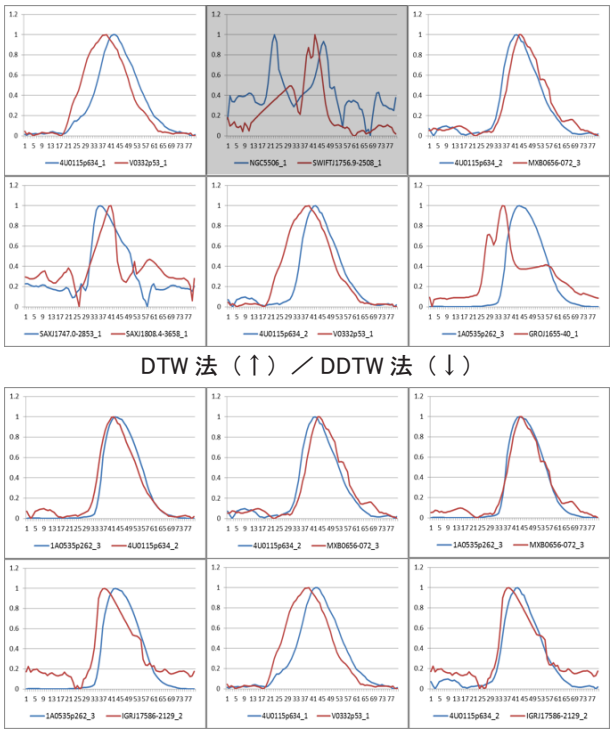
図 8　天体毎のアウトバースト検出数の分布．検出バースト数が突出して多い（29 個）天体は SMCX-1 である．

表 1　DTW 法による類似検索（類似度上位 20 件）

天体1	OB No.	天体2	OB No.	類似度
4U0115p634	1	V0332p53	1	0.018236731
NGC5506	1	SWIFTJ1756.9-2508	1	0.018513513
4U0115p634	2	MXB0656-072	3	0.018897482
SAXJ1747.0-2853	1	SAXJ1808.4-3658	1	0.022549738
4U0115p634	2	V0332p53	1	0.02857227
1A0535p262	3	GRO J1655-40	1	0.029279561
AXJ1749.1-2639	1	H1417-624	1	0.030407902
1A0535p262	2	V0332p53	1	0.030461565
1A0535p262	2	4U0115p634	1	0.03081885
1A0535p262	3	4U0115p634	1	0.032547949
EXO 2030p375	1	V0332p53	1	0.032651616
4U0115p634	1	EXO 2030p375	1	0.033370397
EXO 2030p375	1	H1417-624	1	0.035118511
1A0535p262	3	V0332p53	1	0.035500701
4U0115p634	1	GX304-1	1	0.035695783
1A0535p262	3	GX304-1	1	0.035700165
GX304-1	1	V0332p53	1	0.036771114
1A0535p262	2	4U0115p634	2	0.0392177
GX339-4	2	H1417-624	1	0.041358905
G inga1843p00	1	MXB0656-072	3	0.045164234

表 2　DDTW 法による類似検索（類似度上位 20 件）

天体1	OB No.	天体2	OB No.	類似度
1A0535p262	3	4U0115p634	2	0.002832915
4U0115p634	2	MXB0656-072	3	0.003044645
1A0535p262	3	MXB0656-072	3	0.003189259
1A0535p262	3	IGRJ17586-2129	2	0.00457833
4U0115p634	1	V0332p53	1	0.004600994
4U0115p634	2	IGRJ17586-2129	2	0.005378599
1A0535p262	3	SAXJ1747.0-2853	1	0.00576885
1A0535p262	2	4U0115p634	2	0.006898757
4U0115p634	2	Mrk509	1	0.007081451
4U0115p634	2	SAXJ1747.0-2853	1	0.007212816
IGRJ17586-2129	2	MXB0656-072	3	0.007630367
AXJ1749.1-2639	1	EXO 2030p375	1	0.007875901
1E1743.1-2843	2	IGRJ17586-2129	2	0.008599417
GRO J1655-40	1	MXB0656-072	3	0.009489001
1A0535p262	2	MXB0656-072	3	0.009878791
GX304-1	1	MXB0656-072	3	0.010134271
GX339-4	2	XTEJ1752-223	1	0.010366353
4U0115p634	2	IGRJ17497-2821	1	0.012344659
4U0115p634	1	MXB0656-072	3	0.012853684
1A0535p262	3	GX304-1	1	0.01286813



DTW 法（↑）／DDTW 法（↓）

図 9　DTW 法，DDTW 法による類似検索（左上から右に 1 位から 6 位まで）．

がランキングされている．これは，ある天体に対する観測量が時間的に一定ではなかったり，天体が暗くなることなどによる統計誤差である．このため DTW 法における特異点の不自然な対応付けによって，距離が低く評価されてしまったものと思われる．精度向上のためには，このようなノイズの乗ったデータを事前にフィルタリングする，観測値に付随している観測値に関するメタデータを利用したランキングの補正を行うことなどが考えられる．一方，DDTW 法では，DTW 法のような問題に対応しているため，当該データの順位は低く押さえられている．

滑り窓による類似検索

滑り窓法による類似検索を行ない，結果を評価した．具体的には，それぞれ DTW 法，DDTW 法による距離の閾値を 0.5，0.025 とし，これより小さい距離を示した区間が連続して三つ以上出現したものについて，その平均距離の順に上位 20 件を抽出した（表 3，表 4）．ここで，表における類似部平均は，DTW 距離が閾値を下回った部分区間の DTW 距離の平均である．図 10 は上位 6 件のプロットである．DTW 法の結果については，時系列データ全体の類似度では上位にランクされなかった，部分的な類似区間を含むデータが上位にランクされていることが分かる．これにより，滑り窓による部分的な類似度による類似検索によって，単純な DTW 法（DDTW 法）では検索できないデータを検索することが可能であることが示された．一方，DDTW 法の 4 位，5 位では，一見すると類似していないデータが上位にランキングされていることが分かる．これは，ノイズ等何らかの理由により一定の強度が持続的に維持されているデータ同士に対する結果である．DDTW 法は信号強度の絶対値ではなくその変化率に着目しているため，一定の値を維持している区間について，滑り窓内の類似度が極めて高くなってしまう．DDTW 法では一定の強度が維持されるようなデータを含む場合に，不適切な結果が出力される可能性があることが分かった．これに対しては，滑り窓を適用する区間や幅などを調整すること，一定のレベルが維持されるデータについてはアウトバーストではないと判断して，事前にフィルタリングするなどの対応が必要であると考えられる．

表 3　滑り窓による類似検索（DTW 法，類似度上位 20 件）．

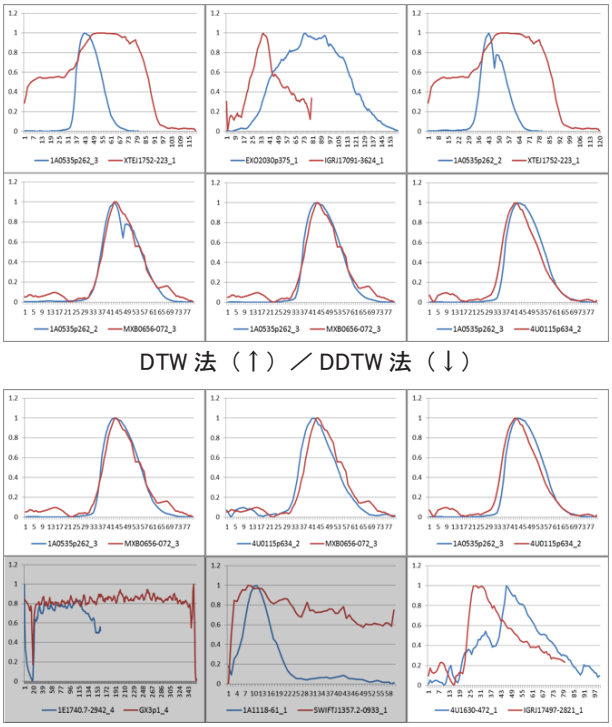
天体1	OB No.	天体2	OB No.	類似部平均
1A0535p262	3	XTEJ1752-223	1	0.009636776
EXO 2030p375	1	IGRJ17091-3624	1	0.029473847
1A0535p262	2	XTEJ1752-223	1	0.032223213
1A0535p262	2	MXB0656-072	3	0.032564217
1A0535p262	3	MXB0656-072	3	0.035630179
1A0535p262	3	4U0115p634	2	0.036466685
1A0535p262	2	4U0115p634	2	0.037292779
4U1702-429	3	XTEJ1701-407	1	0.03775554
1E1743.1-2843	2	4U0115p634	2	0.03797757
EXO 2030p375	1	MXB0656-072	3	0.038530666
1E1743.1-2843	2	MXB0656-072	3	0.039156731
G inga1843p00	2	SWIFTJ1539.2-6227	1	0.04002407
4U0115p634	2	AXJ1744.8-2921	1	0.040474071
4U0115p634	2	MXB0656-072	3	0.044306802
GRO J1655-40	1	SWIFTJ1539.2-6227	1	0.04590276
IGRJ17473-2721	1	IGRJ17586-2129	2	0.04634014
IGRJ17497-2821	1	XTEJ1810-189	1	0.052331623
IGRJ17473-2721	1	XTEJ1810-189	1	0.052364995
EXO 2030p375	2	V0332p53	1	0.05412503
H1417-624	1	SWIFTJ1539.2-6227	1	0.05486193

表 4　滑り窓による類似検索（DDTW 法，類似順上位 20 件）．

天体1	OB No.	天体2	OB No.	類似部平均
1A0535p262	3	MXB0656-072	3	0.001423682
4U0115p634	2	MXB0656-072	3	0.001616227
1A0535p262	3	4U0115p634	2	0.001620353
1E1740.7-2942	4	GX3p1	4	0.002083552
1A1118-61	1	SWIFTJ1357.2-0933	1	0.002356777
4U1630-472	1	IGRJ17497-2821	1	0.002480686
1A0535p262	2	4U0115p634	2	0.002637977
GRO J1655-40	1	XMMUJ054134.7-682550	1	0.003065809
1A1118-61	1	GRO J1655-40	1	0.003101865
1E1740.7-2942	3	IGRJ17473-2721	1	0.003120295
4U0115p634	2	Mrk509	1	0.003164238
4U1630-472	1	XMMUJ054134.7-682550	1	0.003225372
AXJ1749.1-2639	1	SWIFTJ1753.5-0127	1	0.003267231
GRO J1655-40	1	H1417-624	1	0.003394996
4U0115p634	2	IGRJ17586-2129	2	0.003484066
1E1740.7-2942	4	GX9p9	2	0.003490676
GX3p1	4	XTEJ1701-462	2	0.00352943
1E1740.7-2942	4	GRS1724-308	1	0.003600989
1A0535p262	3	1E1743.1-2843	2	0.003705902
1A1118-61	1	XMMUJ054134.7-682550	1	0.003845263

二分割による類似検索

二分割による類似検索を行ない，結果を評価した．具体的には，X 線強度が最大になる点で時系列データを二分割し，前半（後半）のみの部分系列に対して，DTW 法，DDTW 法による距離計算を行なった．ここでは，後半部分のみの結果を示す．表 5，表 6 は上位 20 件の結果，



DTW 法（↑）／DDTW 法（↓）

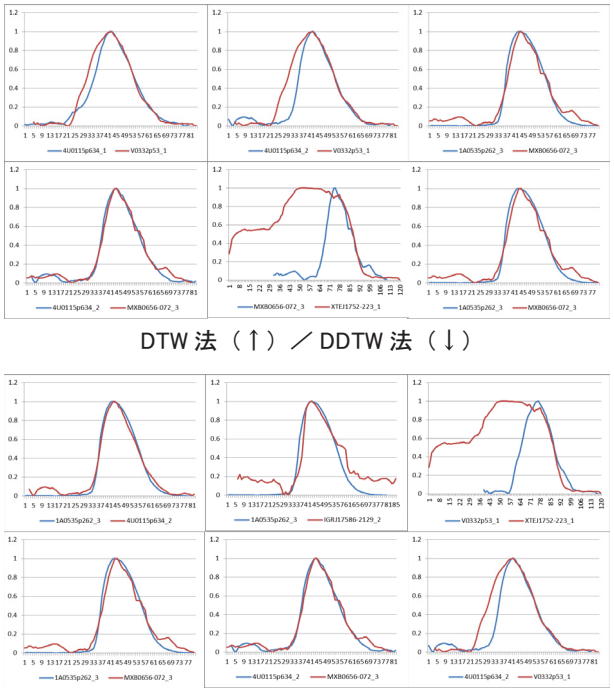
図 10　滑り窓による類似検索（左上から右に 1 位から 6 位まで）．

表 5 二分割による類似検索（DTW 法，類似度上位 20 件）.

天体1	OB No.	天体2	OB No.	類似度
4U0115p634	1	V0332p53	1	0.005059078
4U0115p634	2	V0332p53	1	0.005499665
1A0535p262	3	MXB0656-072	3	0.006590148
4U0115p634	2	MXB0656-072	3	0.00678463
4U0115p634	1	MXB0656-072	3	0.007124647
MXB0656-072	3	XTEJ1752-223	1	0.00728251
1A0535p262	2	MXB0656-072	3	0.007335766
1A0535p262	3	V0332p53	1	0.007446669
MXB0656-072	3	V0332p53	1	0.007546729
AXJ1749.1-2639	1	EXO2030p375	1	0.00769272
1A0535p262	3	XTEJ1752-223	1	0.008206249
1A0535p262	3	1A1118-61	1	0.008858237
Ginga1843p00	2	SAXJ1750.8-2900	2	0.008911676
1A1118-61	1	MXB0656-072	3	0.009128038
1A0535p262	3	AqlX-1	3	0.009276455
1A1118-61	1	4U0115p634	2	0.009444722
GX304-1	1	MXB0656-072	3	0.009608333
1A0535p262	3	GROJ1655-40	1	0.010601421
1A0535p262	3	MAXIJ1409-619	1	0.011048155
1E1743.1-2843	2	GX339-4	2	0.011155077

表 6 二分割による類似検索（DDTW 法，類似度上位 20 件）.

天体1	OB No.	天体2	OB No.	類似度
1A0535p262	3	4U0115p634	2	0.000445267
1A0535p262	3	IGRJ17586-2129	2	0.000762409
V0332p53	1	XTEJ1752-223	1	0.000885903
1A0535p262	3	MXB0656-072	3	0.000950769
4U0115p634	2	MXB0656-072	3	0.000954245
4U0115p634	2	V0332p53	1	0.001163465
1A0535p262	3	1E1743.1-2843	2	0.001166002
1E1740.7-2942	4	MXB0656-072	3	0.001173052
1A0535p262	3	4U0115p634	1	0.001268383
4U0115p634	2	IGRJ17586-2129	2	0.001271939
1A0535p262	3	V0332p53	1	0.001277466
1A0535p262	3	XTEJ1752-223	1	0.001315496
1E1743.1-2843	2	IGRJ17586-2129	2	0.001364595
4U0115p634	1	V0332p53	1	0.001400574
4U0115p634	1	XTEJ1752-223	1	0.001469094
1E1740.7-2942	4	4U0115p634	2	0.001555429
4U0115p634	2	GX1p4	2	0.001619554
4U1702-429	4	EXO2030p375	2	0.001665785
IGRJ17586-2129	2	XTEJ1946p274	3	0.001814627
4U0115p634	2	XTEJ1946p274	3	0.001837686



DTW 法（↑）／DDTW 法（↓）

図 11 二分割による類似検索（左上から右に 1 位から 6 位まで）.

図 11 は上位 6 件のプロットである. 図から分かるように，DTW 法，DDTW 法とも，最大値からの値の下がり方が類似しているデータを検索できていることが分かる.

5 まとめ

本研究では，X 線アウトバーストの時系列データの類似検索手法を提案した．時系列データの検索手法である DTW 法，DDTW 法をベースとして，時系列データの部分的な類似性を考慮に入れた検索を行なうために，滑り窓を用いた方法と時系列データを X 線強度が最大になる点で前後に二分割し，部分データに対して距離計算を行なう手法を提案した．また，実験による評価によって，提案手法の有効性を検証した．今後は，アウトバースト部分の検出精度を向上し，ノイズの乗ったデータに対する適切な処理を行うことによって，類似検索自体の精度向上を目指す．また，既存のデータに対し網羅的に類似検索を行ない，新たな知見の獲得に取り組む．さらに，提案した手法を Web サービス化して公開することも検討する予定である.

参考文献

1) Satoshi Nakahira and MAXI Team. Black hole novae observed with MAXI. *The Astronomical Herald*, Vol. 103, No. 3, pp. 166–175, 2012.

2) M.Matsuoka. et al. The MAXI mission on the ISS: Sciences and instruments for monitoring all sky Xray images. *PASJ*, Vol. 61, No. 999, 2009.

3) Davood Rafiei and Alberto Mendelzon. Similaritybased queries for time series data. In *Proc. 1997 ACM SIGMOD international conference on Management of data (SIGMOD'97)*, 1997.

4) I. Popivanov. Similarity search over time-series data using wavelets. In *Proc. 18th International Conference on Data Engineering (ICDE2002)*, pp. 212–221, 2002.

5) Kiyoung Yang and Cyrus Shahab. A PCA-based similarity measure for multivariate time series. In Proc. 2nd ACM international workshop on Multimedia databases (MMDB'04), pp. 65–74, 2004.

6) H. Sakoe and S. Chiba. A dynamic programming approach to continuous speech recognition. In *Proc. 7th International Congress on Acoustics*, pp. 65–69, 1971.

7) Eamonn J. Keogh and Michael J. Pazzani. Derivative

dynamic time warping. In *Proc. 1st SIAM International Conference on Data Mining (SDM2001)*, 2001.

8) Jon Kleinberg and ´ va Tardos. *Algorithm Design*. Addison Wesley, 2005.

9) Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proc. KDD Workshop*, pp. 359–370, 1994.

10) E. G. Caiani, A. Porta, G. Turiel, M. Muzzupappa, S. Pieruzzi, F. Grema, C. Malliani, A. Cerutti, and S. Cerutti. Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. In *IEEE Computers in Cardiology*, Vol. 25, 1998.

11) 石川雅弘 , 吉川昂伯 , 陳漢雄 , 古瀬一隆 , 大保信夫 . 類似部分区間検索のためのタイムワーピング距離の下限値計算 . *DBSJ Letters*, Vol. 6, No. 1, pp. 25–28, 2007.

12) F. Itakura. Minimum prediction residual principle applied to speech recognition. *Audio*, Vol. 23, No. 1, pp. 67–72, 1975.

13) H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. 26, No. 1, pp. 43–49, 1978.

14) J. B. Kruskal and M. Liberman. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, chapter* The Symmetric Time-Warping Problem: from Continuous to Discrete, pp. 125–161. Addison-Wesley, 1983.

15) Stan Salvador and Philip Chan. FastDTW: Toward accurate dynamic time warping in linear time and space. *Time*, Vol. 11, No. 5, pp. 70–80, 2004.

16) fastdtw: Dynamic time warping (DTW) with a linear time and memory complexity. `http://code.google.com/p/fastdtw/`.

17) NASA. Swift/BAT hard X-ray transient monitor. `http://swift.gsfc.nasa.gov/docs/swift/results/transients/`.