

Smith Waterman 法を利用した fNIRS データの類似部分抽出システムの提案

西井 琢真

Takuma NISHII

1 はじめに

近年、脳血流の変化を測定することにより非侵襲的に脳機能マッピングを行う fNIRS(functional Near-Infrared Spectroscopy) や fMRI(functional Magnetic Resonance Imaging) が注目を集めている¹⁾。これらの装置は、脳機能の解明に役立ち、種々の病理の判定や生体信号によるコンピュータ操作などに利用されている²⁾。しかし、これらの装置の性能が向上し出力される時系列データ量が增大すると解析者が効率的にデータを解析できないという問題も生じる。効率的にデータを処理するための課題はいくつか存在するが、その 1 つに解析者がどのデータのどの部分に着目すれば良いのかという問題がある。

本研究ではこの問題を解決するために複数の時系列データの中から特徴的な部分を抽出するアルゴリズムを提案し、システムを構築することで解析者の負担を軽減する。

具体的には、Smith Waterman 法^{3) 4)}を利用した fNIRS データの類似部分抽出アルゴリズムの提案し、それを実装したシステムの評価を行う。Smith Waterman 法を利用した fNIRS データの類似部分抽出アルゴリズムは、fNIRS の出力データとする時系列データに対して、時系列データの再量子化を行ない、Smith Waterman 法を適用することで、時系列データからの類似部分を抽出する。fNIRS は出力データが大量であるため、解析に用いられるアルゴリズムは高速である必要がある。そのため、時系列データからの類似部分の抽出については、バイオインフォマティクスの分野で盛んに扱われる Smith Waterman 法に注目した。Smith Waterman 法は、アルゴリズムの並列性が高く、CPU マルチスレッドプログラミングや GPU を用いた高速な探索を期待できる。

2 関連研究

脳機能イメージング装置の解析ソフトとしては、日立メディコ社製の ETG-7100 における Wave Analysis ソフト⁵⁾ やスペクトラテック社製の Spectratech OEG-16 の解析ソフトなど装置に付属するソフトが挙げられる。これらのソフトを用いることで生データのグラフ化や簡単な処理は可能である。また、Matlab を用いた解析ソフト「NIRS-SPM」⁶⁾ や Source Signal Imaging 社製の脳波解析プログラム「EMSE」⁷⁾ を用いることでより高度な波形処理や脳の 3D モデリングが可能である。しかしながら、解析すべきデータのうち、どこに着目すべきかを提示する事を目的としたソフトはほとんどないといえる。

また、本論文では出力データを高速に処理するメソッ

ドとして Smith Waterman 法を用いて、そのパラメータ調整を行った。SW 法はアルゴリズムの並列性が高く、GPU を用いた高速実行のための様々な実装が試みられている^{8) 9) 10) 11)}。

3 同源性検索と SW(Smith Waterman) 法

同源性検索は、バイオインフォマティクスの分野で広く用いられている文字列検索アルゴリズムであり、DNA や塩基配列の類似度測定や類似部分の抽出が可能である。SW 法は動的計画法の 1 種であり、全ての部分文字列の比較を行うことで類似部分を最適化する。長さが m と n の文字列から類似部分を抽出する場合、アルゴリズムのオーダーは $O(mn)$ である。



Fig.1 String sequence alignment using SW

4 SW 法を用いた 2 つの時系列データの類似部分の抽出方法

本論文では時系列データの再量子化と同源性検索の組み合わせによる、2 つの時系列データからの類似部分の抽出手法を用いた¹²⁾。Fig.2 に抽出手法の流れを示す。まず、時系列データを再量子化する。本論文では再量子化は文字列化を意味する。例えば、Fig.2 の上部の時系列データは“BAABCCB”に、下部の時系列データは“ABCCBAA”に変換される。再量子化の手法としては、SAX(Symbolic Aggregation approxImation) や等間隔領域分割がある¹³⁾。時系列データを文字列に変換することによって、同源性検索を適用することが可能となる。この手法により“BAABCCB”と“ABCCBAA”から類似部分として“ABCCB”を取り出すことができる。このように時系列データの再量子化と同源性検索による部分文字列の抽出によって時系列データの類似部分の抽出が可能になる。本論文では、本手法を用いて抽出された部分を類似部分として定義した。fNIRS の出力データに対して提案手法を適用した例を Fig.3(a), Fig.4(b) に示す。

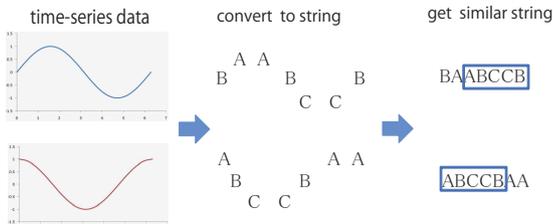


Fig.2 The outline of proposed method

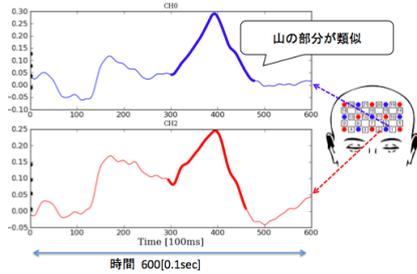


Fig.3 Sample extracted data using the proposed algorithm (Time phase shifts in the two data are the same)

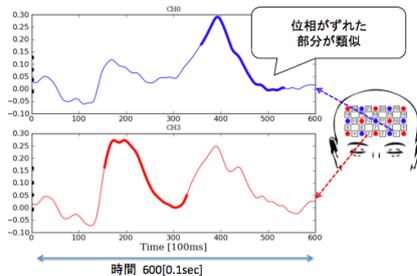


Fig.4 Sample extracted data using the proposed algorithm (Time Phase shifts in the two data are the different)

5 解析システムの提案

5.1 概要

このシステムの目的は、データ解析において従来は注目されなかった脳部位データのどれに着目すべきかを提示することで解析者に新たな知見を与えることである。現在、fNIRSの解析者は注目した脳部位のデータに対してのみデータ処理を行っており、その方法では注目している脳部位以外に重要な要素がある場合にそれを見落としてしまう可能性がある。これを解決するためには、ある特徴的な波形に類似した波形が他の脳部位のどこにあるのかを特定することが有効なのではないかといえる。これにより、解析者は低負担で着目すべき部位を見つけることが可能である。

5.2 機能

ここでは作成するシステムの機能について説明する。システムに求められる機能は以下の3つである。

機能1: fNIRSのファイルを読み込み、CHデータ一覧を表示

機能2: CHを1つ選択し、そのCHデータに類似するデータを表示

機能3: あるCHデータの範囲を指定し、それと類似するものを他のCHデータ群から表示

6 SW法の処理速度性能

6.1 概要

第5章で述べた解析システムの機能を実現するためには、提案手法を何度も繰り返す必要がある。この手法において、特に処理に時間がかかるのはSW法の部分であると予想される。そのため、システムを実現するための処理時間を計測する必要がある。一般的によく用いられるシングルCPU、シングルスレッドのアルゴリズムでは、処理の反応が返ってくるまでに十数秒の時間を要するという問題がある。解析者が快適に作業するためには、クリックなどのアクションをしてから約1~2秒以内で反応が返ってくることを望ましいと考えている。そのため、システムを実現するための処理時間を計測し、マルチコアCPU、マルチスレッドを利用してSW法を高速化することを検討する。

6.2 実験目的

ここでは、マルチコアCPU環境においてマルチスレッドを利用したSW法の速度計測を行った。マルチスレッドのSW法を実行するためには、スレッド数と部分ブロックのサイズをパラメータとして指定する必要がある。パラメータの設定によっては高速化がうまくいかない場合も考えられるため、事前に最適なパラメータを把握する必要がある。また、解析システムにおいて実際に機能2、3がどれくらいの処理時間になるかを計測し、高速なレスポンスを実現できるかどうかを把握する必要がある。

6.3 実験結果

文字列の長さを256, 512, 1024, 2048, 4096と変化させSW法の処理時間を計測した。それぞれの実行計測は3回ずつ行い、実験結果にはその中央値を用いた。それぞれの文字列サイズにおける最適パラメータをTable.2に示す。使用したマシンの環境をTable.3に示す。紙面の都合上、Fig.8のみ結果を示す。fNIRSのサンプリングレートは0.1秒であるため、1秒あたり10文字となる計算である。そのため、256~4096文字は、約25秒~410秒の実験時間に該当する。インターフェースの機能1~3に必要なSW法の繰り返し回数と、文字列の長さ600のときの処理時間をTable.1示す。なお、文字列の内容によって処理時間が変化することはない。

6.4 評価

実験の結果、パラメータの設定によっては処理時間が数倍異なることが分かった。このことから処理の高速化のためには、パラメータ調整も必要であるといえる。しかし、Table.2にある最適パラメータはマシン

Table1 The number of times of executing SW algorithm for functions (1) - (3)

機能	SW 法の繰り返し回数	処理時間 [sec]
機能 1	0	0
機能 2	24	0.165
機能 3	24	0.165

Table2 Optimum parameters for each table size

文字列サイズ	スレッド数	サブブロックサイズ
256	4	32
512	4	32
1024	4	128
2048	8	32
4096	8	128

Table3 Operating environment

OS	Ubuntu10.10 64bit
CPU	Intel Core i7-2600 4cores 3.40GHz
RAM	8GB

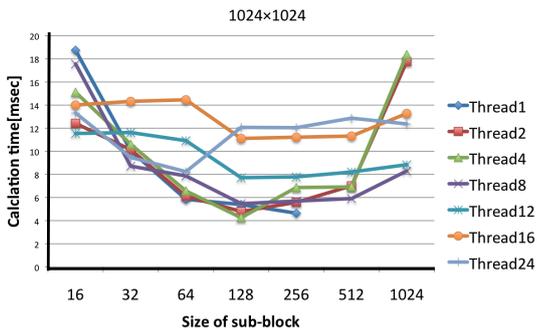


Fig.8 Calculation time with along to the number of threads (table size 1024 × 1024)

環境により変化することが考えられる。また、Table.1 より、SW 法の処理時間は機能 2, 3 を実行すると 0.165 秒となり、1 秒以内の高速なレスポンスが可能であることが分かった。

7 構築したシステムによる fNIRS データの解析

7.1 概要

解析システムを用いて、実際の fNIRS 出力データを解析した。用いた実験データは、タスクが「ストループ効果に関するタスク」、実験時間は 60sec、サンプリングレートは 0.1sec である。この場合、文字列の長さが 600 になることから、Table.2 より、スレッド数を 4、サブブロックサイズを 32 にすることが適切だと考えた。SW 法の処

理時間は、Table.1 と同様となった。

7.2 ストループ効果

ストループ効果とは、色と語の意味が不一致な単語に対して、色命名反応がなされるとき、反応が困難になり、反応時間が増大するという認知的葛藤現象である^{14) 15)}。先行研究により¹⁶⁾、ストループ効果において左半球の下前頭回付近での反応が認められている。反応が確認されている左半球におけるデータを特徴のあるデータと捉え、それと類似するものを見つけられれば解析者の助けになると考えられる。

7.3 結果

解析システムの処理結果を Fig.9 に示す。Fig.9 を見ると、反応が確認されている左半球のある CH と類似した波形が右半球、前頭部にも現れていることがわかる。このように注目部位以外に観るべき点を示すことで解析に役立つと考えられる。

8 まとめと今後の課題

本稿では、大量の実験データを出力する fNIRS のデータ解析のためのアルゴリズムとシステムの提案と評価を行った。相同性検索を用いて fNIRS データから類似部分を抽出するアルゴリズムを紹介し、それを実現したシステムの機能とレスポンスについて述べた。その結果、現在想定している機能についてシステムのレスポンスは十分高速である事がわかった。また、実際の fNIRS の出力データに対してシステムを適用し、類似部分がどのように表示され、解析者の負担軽減に繋がるかについて述べた。

今回作成したインターフェースは、ユーザが基準となる CH を選択し、それに類似したデータを表示するというものであった。今後は、基準となる CH の自動選択機能も検討したい。

また、今回は提案手法を用いて抽出された部分を類似部分として定義したが、今後は時系列データの変化の度合い（山の数）を類似の定義として用いることを考えている。そのため、時系列データの山の数を意識した前処理が重要になってくると考えている。

参考文献

- 1) James C. Eliassen; Erin L. Boespflug; Martine Lamy; Jane Allendorfer; Wen-Jang Chu; Jerzy P. Szaflarski. Brain-Mapping Techniques for Evaluating Poststroke Recovery and Rehabilitation. *Neuroplasticity: Changing Minds and Changing Function*.
- 2) Xu Cuia; Signe Braya; Daniel M. Bryanta; Gary H. Gloverc; Allan L. Reissa. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*.
- 3) T F Smith; M S Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, Vol. 147, No. 1, pp. 195–197, March 1981.

- 4) Hideyuki Takagi. Implementation of the smith-waterman algorithm on a reconfigurable super-computing platform. HPRCTA '07 Proceedings of the 1st international workshop on High-performance reconfigurable computing technology and applications, 2007.
- 5) 株式会社日立メディコ. Optical topography etg-7100. <http://www.hitachi-medical.co.jp/product/opt/etg/func.html>.
- 6) BiSPL(Bio Imaging Signal Processing Lab). Nirs-spm. <http://bisp.kaist.ac.kr/NIRS-SPM.html>.
- 7) Source Signal Imaging Inc. Emse 脳波解析プログラム. <http://www.miyuki-net.co.jp/jp/product/emse.htm>.
- 8) Keisuke Dohi; Khaled Benkrid; Cheng Ling. Highly efficient mapping of the smith-waterman algorithm on cuda- compatible gpus. ASAP' 10, Vol. 36, pp. 29–36, 2010.
- 9) Yongchao Liu; Douglas L Maskell; Bertil Schmidt. Cudasw++: optimizing smith-waterman sequence database searches for cuda-enabled graphics processing units. *BMC Research Notes*, 2009.
- 10) Svetlin A Manavski; Giorgio Valle. Cuda compatible gpu cards as efficient hardware accelerators for smith-waterman sequence alignment. *BMC Bioinformatics*, 2008.
- 11) nVIDIA. Tesla bio workingbench. http://www.nvidia.co.jp/object/tesla_bio_workbench_jp.html.
- 12) Takuma Nishii; Tomoyuki Hiroyasu; Masato Yoshimi; Mitsunori Miki; Hisatake Yokouchi. Similar subsequence retrieval from two time series data using homology search. *Systems Man and Cybernetics*, pp. 1062–1067, 2010.
- 13) 廣安知之, 西井琢真, 吉見真聡, 三木光範, 横内久猛. 相同性検索を用いた 2 つの時系列データからの類似部分抽出手法と DTW による類似部分の評価. 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol. 2010, No. 24, pp. 1–6, 2010-09-21.
- 14) 嶋田博行. ストループ効果-認知心理学からのアプローチ. 培風館, 1994.
- 15) Stroop J.R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, pp. 643–661, 1935.
- 16) A. Ehlis et al. Multi-channel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent stroop trials. *Biological psychology*, pp. 315–331, 2005.

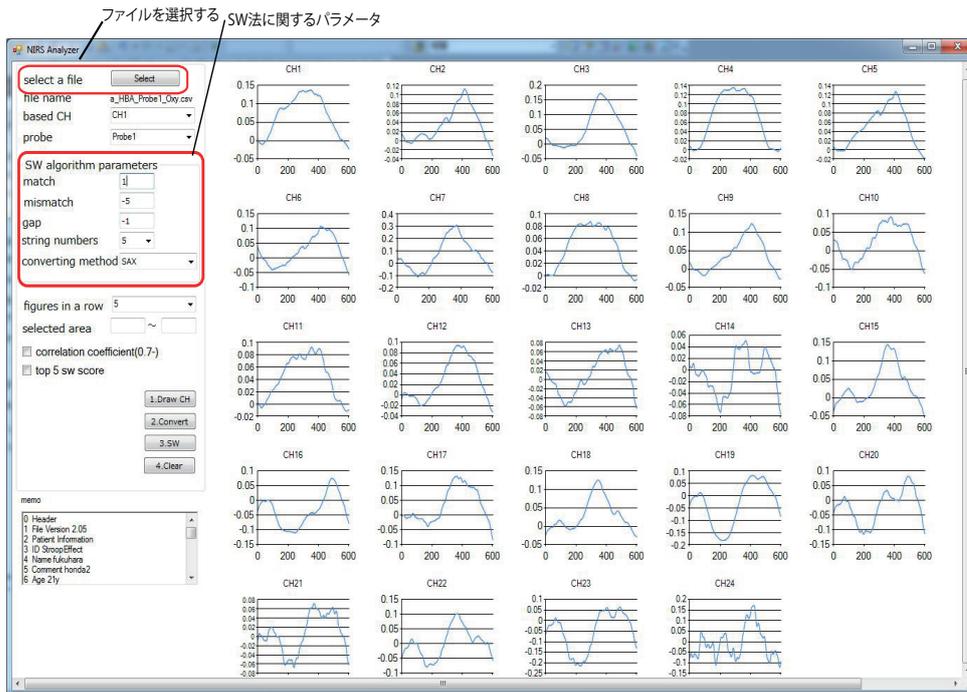


Fig.5 Reading fNIRS output data and showing all the CHs data

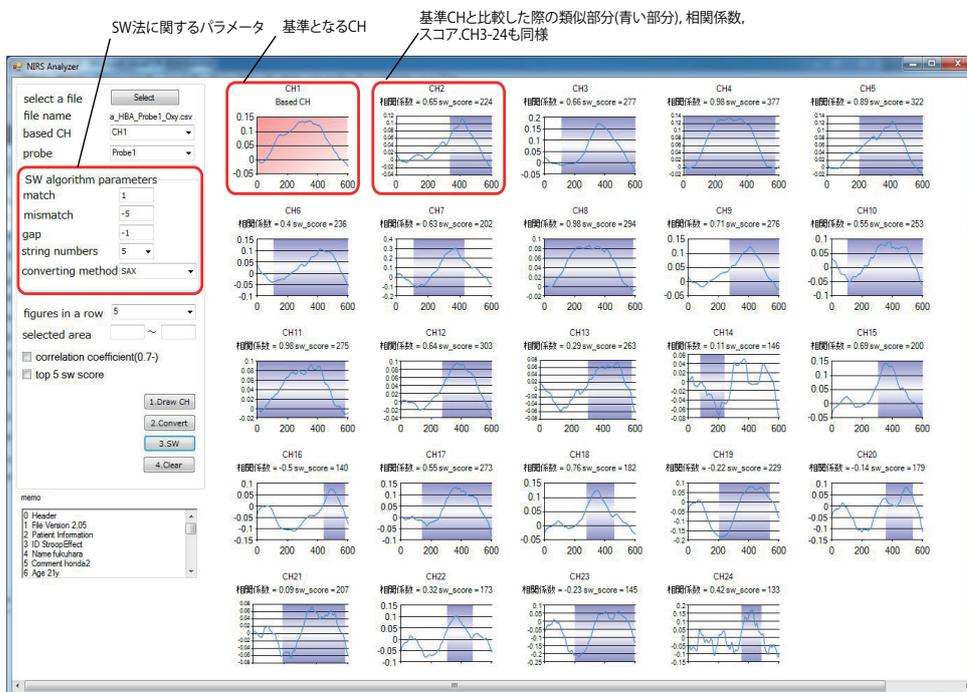


Fig.6 Selecting a CH and then showing the related data

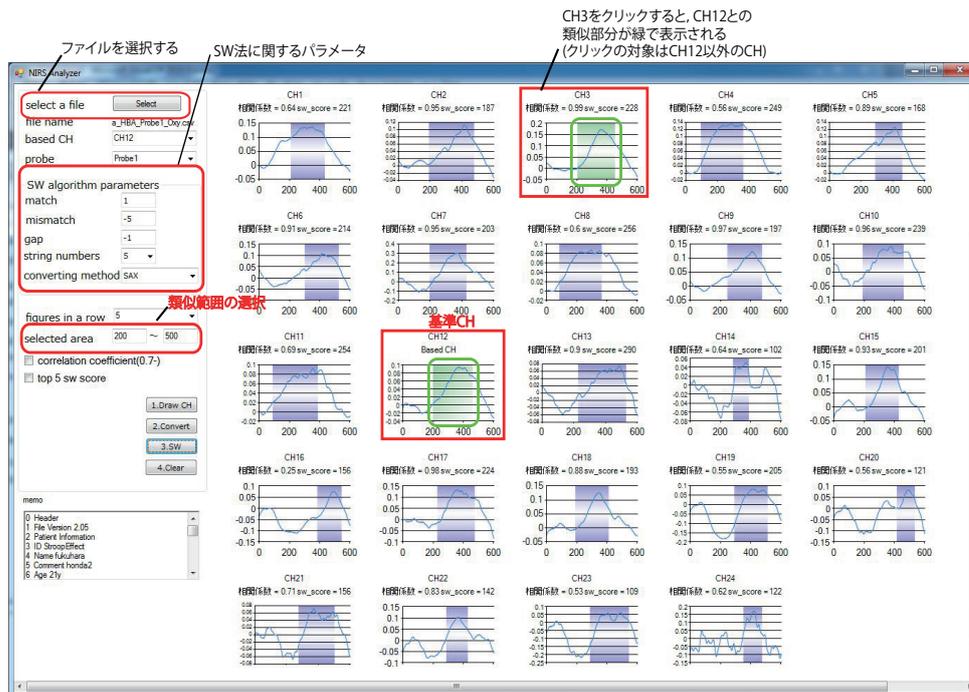


Fig.7 Selecting a CH and its part and then showing the related data

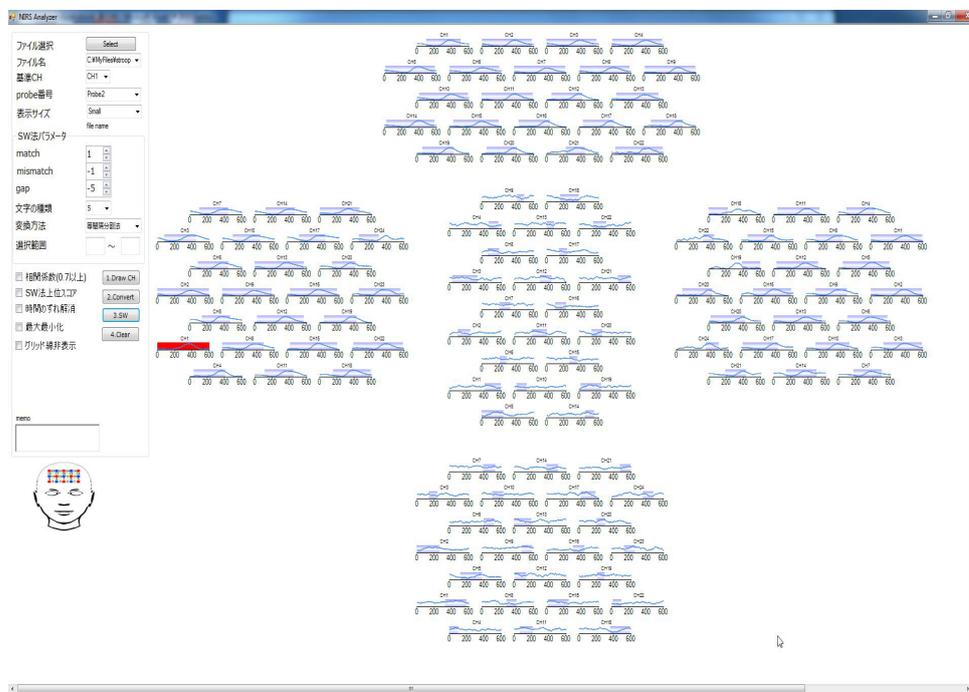


Fig.9 Output data of fNIRS(CH1 of Probe2 and its similar parts)