1B2-OS-02a-1

持続可能なコンセプトドリフト適応型モニタリングシステムの提案

Proposition of Sustainable and Concept Drift-Adaptive Monitoring System

坂本 悠輔[*1] Yusuke Sakamoto　福井 健一[*2] Kenichi Fukui　Daniela Nicklas[*3]　森山 甲一[*2] Koichi Moriyama　沼尾正行[*2] Masayuki Numao

[*1] 大阪大学大学院情報科学研究科 Graduate School of Information Science and Technology, Osaka University

[*2] 大阪大学産業科学研究所 The Institute of Scientific and Industrial Research, Osaka University

[3*]Carl Von Ossietzky University Oledenburg, Germany

We propose a monitoring system based on concepts composed of a number of features as a monitoring target. Self-Organizing-Maps (SOM) is utilized to obtain concepts and to classify a new event into one of the concept. To make the system sustainable, the system has to deal with concept drift such as change of a concept and appearance of a new concept over time. In this work, we tried to validate use of cluster assignment error as a detection criterion of concept drift. We aim to adapt our system for monitoring of damages in a fuel cell.

1. はじめに

現在，機器や装置のモニタリングシステムは危険や異常の検知のために様々な場面で用いられている。例えば温度，電圧などの物理量をシステムの判断基準とするモニタリングシステムが広く用いられている．しかし，この物理量型モニタリングシステムでは，危険や異常の判断は，判断基準としている物理量が閾値を超えるか超えないかという基準で行われるために，閾値が適切に設定されていない場合に危険や異常を正しく検知できない問題や，危険や異常を正しく検知出来たとしても，原因の特定を迅速に行えない問題がある．原因の特定を迅速に行えない問題は，システムの警告があればその都度，物理量や使用状況から使用者が原因を探る必要があるために起こる．

このような問題を解決する方式として，学習ベースモニタリングシステムが考えられる．事前に学習により対象のモデルを構築することで，複数の物理量と故障や異常とを概念（例えば、部品 A の破損など）として関連付けることができる．システムの判断基準を概念とする最大の長所は、システムが異常や危険を検知した際に，一見して異常や危険の原因までわかるということである．しかし，学習ベースの方式では，モニタリング対象にこれまで起きていない故障や劣化が生じると，モデルに含まれない未知の概念に対応できず持続可能性に欠くという問題がある．すなわち，コンセプトドリフト[Zliobaite 10]（より正確には対象 S の観測量の確率分布 Pr(S)の変化や，新たな確率分布 Pr(S’)の出現を指す）に対応する必要がある．ここで，全ての概念を事前に学習しておくことは非現実的であるため，モニタリングと共にモデルの更新が必要となる．一方，逐次入力されるデータからコンセプトドリフトの検出[Gama 04]や，モデルを更新する逐次学習方式[Silva 13]なども提案されているが，これらの研究は学習の観点で検出精度やモデルの追従精度を検証するに留まっており，モニタリングに応用した例はみられない．また，学習にはクラスタリングであればクラスタリング結果の意味付けや，教師あり分類学習であれば教師となるラベルを観測対象に精通した人に付与してもらう必要があること，さらに学習に要する計算時間は一般に事象の入力頻度に比べて長いことが挙げられる．

そこで本研究では，ユーザとのインタラクションが必要で低速な学習機能と，オンラインでの高速な処理を実現するデータストリーム管理システム(DSMS)を組み合わせるアーキテクチャを提案した．本方式では、統計に基づいた動的な閾値管理手法[Gama 04]によりコンセプトドリフトを検出し，適切なタイミングでモデルの更新を行うことで，持続的なモニタリングを目指す．そして初期的な検討として，木材の破損時の音データに対して人工的にコンセプトドリフトを設定した時の，クラスタ割当て誤差の変化を示すことにより，提案アーキテクチャに適したコンセプトドリフト検出法の検討につなげる．

連絡先：坂本悠輔，大阪大学大学院情報科学研究科，〒567-0047 大阪府茨木市美穂ヶ丘 8-1，TEL:06-6879-8426，FAX:06-6879-8428,mail:sakamoto@ai.sanken.osaka-u.ac.jp

2. システム実現のための提案アーキテクチャ

2.1 全体像

提案するアーキテクチャを図 1 に示す．赤枠で示されている処理は，各事象がセンサで観測される度に，高頻度に処理が必要な部分である．これらは，高速に処理する必要があるため，全てオンメモリで処理を行う DSMS 上に実装する必要がある．ここで，DSMS は Nicklas らが提案する汎用フレームワークである Odysseus[Appelrath 12]を利用することで，容易に実現できる．

一方，青枠で示される処理は，観測対象に何らかの変化が起きてコンセプトドリフトが検出された場合にのみ処理を行う．この部分には，ユーザとのインタラクションや学習モデルの更新，アーカイブへの書き出しなどの低頻度処理が含まれる．それぞれの中身について，我々を中心としてこれまで研究を続けてきた燃料電池の損傷評価[福井 10][北川 11]を例に説明する．

先行研究では，燃料電池の損傷の計測には Acoustic Emission(AE)法を用いている．AE 法とは，材料の破壊の際に発生する微弱な弾性波を波形信号として測定することによって，非破壊に損傷を評価する方法である[佐藤 05]．そして，多様な損傷に由来する AE 事象の分類と可視化のためにニューラルネットワークの一種である自己組織化マップ（SOM）を用いている．

2.2 処理の流れ

まず常時計測されるノイズを含む AE センサの信号から，損傷に起因する AE 事象部分を切り出し，FFT による周波数変換を行い，入力ベクトルを得る (Preprocessing)．次に、新規 AE 事象の入力ベクトルを学習済み SOM マップに割り当てる(Cluster Assignment)．ここで，Concept Drift Detection を行い，ドリフトが

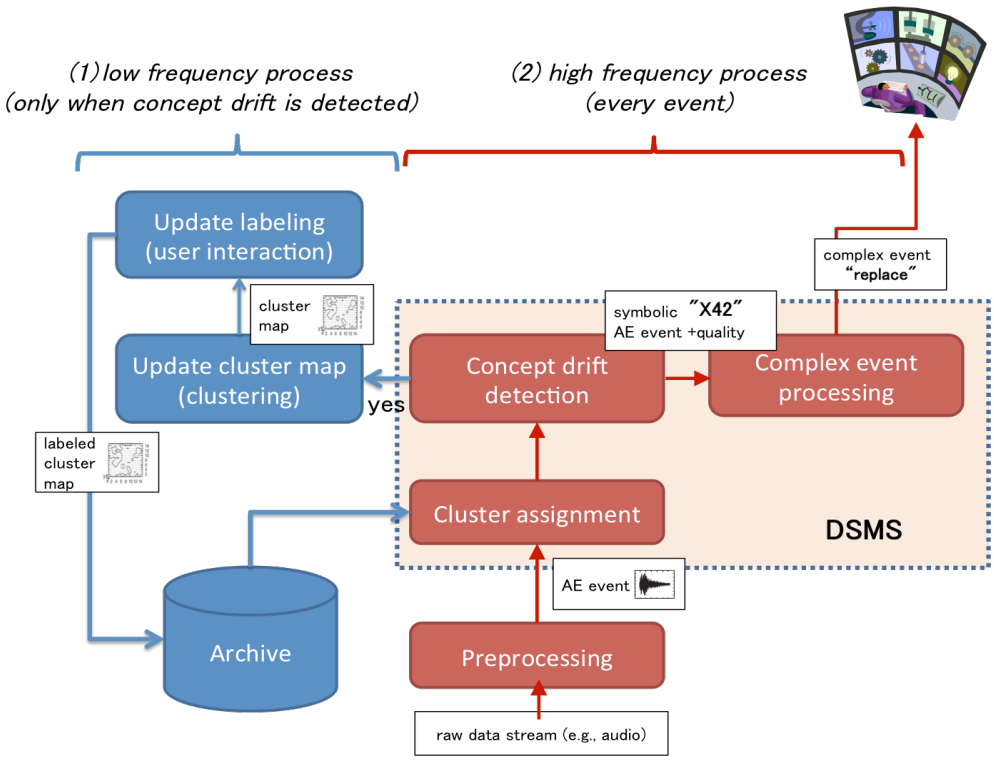

図 1. 提案アーキテクチャ

検出された場合は学習モデルを更新する必要があるので，low frequency process に進み，SOM 学習モデルの更新(Update cluster map)とユーザによるラベリング(Update labeling)を行い，新たなマップを DSMS に受け渡す．ドリフトが検出されなければ，モニタリングの処理として事前に定義したルールベースと照らし合わせて合致すれば警告を出したりアーカイブしたりする(Complex Processing).

2.3 SOM

SOM の特徴は高次元データを教師なしでクラスタリングし，2 次元平面上に視覚的に表現できる点である．SOM は位相空間上に規則的に配置された複数のニューロンノードから構成される．各ノードは入力ベクトルと同次元の参照ベクトルを持っており，SOM に入力ベクトルが与えられると，入力ベクトルと最も近い参照ベクトルを持つノードが勝者ノードとなる．この時，位相空間上で勝者ノードに近いノードほど入力ベクトルに対して強く学習する権利を獲得し，その強さに応じて参照ベクトルを入力ベクトルへと近づけるように学習する．なお、本稿では音波事象の距離測度として周波数スペクトル分布の Kullback-Leibler 情報量を用いるため、カーネル化した SOM[Andras 02] [Boulet 08] を用いた．

2.4 Cluster Assignment

新規 AE 事象は，参照ベクトルとの量子化誤差が最小となるニューロンへ割り当てられる．ここで，x_nを新規 AE 事象の入力ベクトルとすると，x_nが割り当てられる勝者ニューロンの量子化誤差は次式で与えられる．

$$ERROR = \min_{i=1,\dots,m}\left\{K(x_n,x_n) - 2\gamma\sum_j h_{i,C(j)}K(x_n,x_j) + \gamma^2\sum_k\sum_l h_{C(k),i}h_{C(l),i}K(x_k,x_l)\right\} \quad (1)$$

ここで，i:SOM のノード番号，m:ノード数，j,k,lのΣは SOM 学習モデルの作成に用いた学習データについての総和である．$K(x_i,x_j)$はカーネル関数であり，本稿では次式で示される確率分布間の近さを表す Kullback-Leibler 情報量に基づく尺度に基づく KL カーネルを用いた．

$$K(x_i,x_j) = \exp\left(-\beta JS(x_i,x_j)\right) \quad (2)$$

$$JS(x_i,x_j) = KL(x_i,x_j) + KL(x_j,x_i) = \sum_{k=1}^{v}\left\{\frac{x_{i,k}\log x_{i,k}}{x_{j,k}} + \frac{x_{j,k}\log x_{j,k}}{x_{i,k}}\right\} \quad (3)$$

また，$h_{i,j}$は位相空間上のユークリッド距離に基づく近傍関数である．本研究では，ガウス関数を用いた．γは正規化項であり，$\gamma = \frac{1}{\sum_n h_{C(n),i}}$によって与えられる．C(j)は式(1)で計算される勝者ニューロンの番号を表している．

2.5 Concept Drift Detection

本稿では，コンセプトドリフトの検出基準として，SOM による学習モデル，すなわちマイクロクラスタへの割り当て誤差の検討を行う．ここで、検出基準の要件として，学習済みモデルに含まれる概念およびモデルに含まれない概念から得られた新規データの両評価値において，統計的に有意な差が得られる必要がある．さらに，検出法の要件としては，より少ない観測回数で尚且つ 1 次オーダーの計算量で検出する必要がある．そのような検出法として，我々は Gama らの Concept Drift Detection 法[Gama 04]に着目している．ただし，Gama らの研究では，教師

あり学習器における滑走窓中の損失関数値を検出基準としている.

3. 実験

我々は, DSMS 上でのコンセプトドリフト検出を目指し, 実験を行った.

[実験目的]

コンセプトドリフト時に, SOM のマイクロクラスタへの割り当て誤差に優位な差は認められるか確認する.

[データ]

実験では, 木材破損時(3mm×3mm バルサ材、5mm×5mm バルサ材, 竹材, 割箸, シナ材)の音波データを対象データとした. また, 実験簡単化のため, DSMS 上ではなく, SOM のマイクロクラスタへの割り当てを逐次的に行うことで疑似オンライン環境として実験を行った.

実験の処理対象となる音波データの例として, 3mm×3mm バルサ材の波形データと周波数パワースペクトルをそれぞれ図 2 および図 3 に示す.

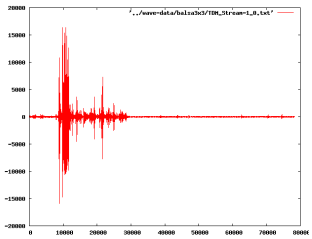

図 2. 3mm×3mm バルサ材の波形データ例

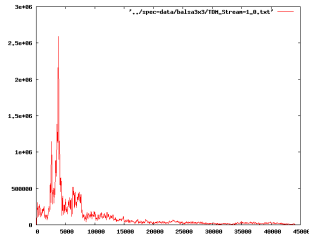

図 3. 3mm×3mm バルサ材の周波数パワースペクトルの例

[実験手順]

1. 3×3 バルサ材(20 個)をトレーニングデータとして, SOM の学習モデルを作成した.
2. 3×3 バルサ材(5 個), 5×5 バルサ材(25 個), 竹材(25 個), 割箸(25 個), シナ材(25 個)から, 以下のようにして人工的なコンセプトドリフトを設定したテストデータセット A), B), C), D)を用意した.
   - A) 3×3 バルサ材(5 個)→5×5 バルサ材(25 個)の順にデータを並べた.
   - B) 3×3 バルサ材(5 個)→竹材(25 個)の順にデータを並べた.
   - C) 3×3 バルサ材(5 個)→割箸(25 個)の順にデータを並べた.
   - D) 3×3 バルサ材(5 個)→シナ材(25 個)の順にデータを並べた.

   なお, A), B), C), D)における 3×3 バルサ材(5 個)データの並びは同じ並びとした.
3. データセット A), B), C), D)それぞれにおいて、1.で作成した学習モデルにデータを逐次的に割り当て, 式(1)で求められる割り当て誤差の変化を観察した.

[実験結果]

データセット A), B), C), D)の実験結果としてそれぞれ図 4, 5, 6, 7 を得た. 図中の横軸は割り当てたデータの入力順を示し, 縦軸はそのときの割り当て誤差を表す. また, すべてのデータセットで最初の 5 個のデータは 3×3 バルサ材で, Data Number 4(図中の青線)を境にデータの種類が変化する. 例えば, データセット A)ならば 3×3 バルサ材→5×5 バルサ材. よって図中の青線を境に人工的なコンセプトドリフトが設定されている.

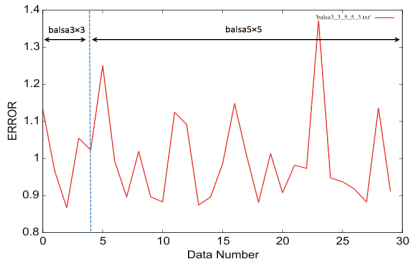

図 4. データセット A の割り当て誤差の変化

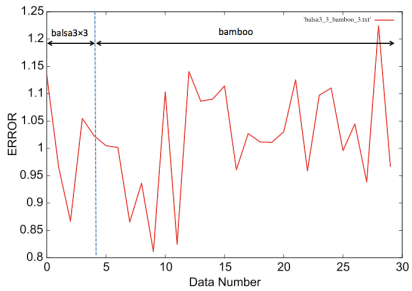

図 5. データセット B の割り当て誤差の変化

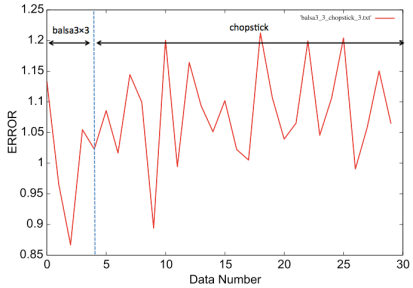

図 6. データセット C の割り当て誤差の変化

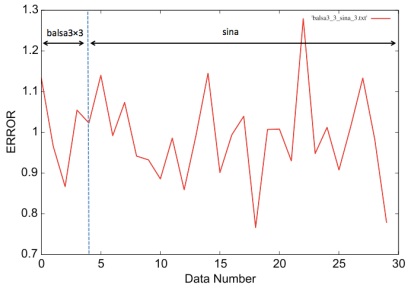

図 7. データセット D の割り当て誤差の変化

以上より， 今回の実験からは，青の破線で示した人工的なコンセプトドリフトを境にして， SOM のマイクロクラスタへの割り当て誤差に当初期待していた優位な差は認められなかった．

[考察]

SOM によるクラスタ割当て誤差では，顕著な差が得られなかった原因として，学習データと同じ種類のテストデータの少なさ(3×3 バルサ材 5 個のみ)，また実験に用いた音の周波数スペクトルの分布にそもそも顕著な差がない可能性，もしくはクラスタ割当て誤差の基準が不適切であったことが考えられる．まず，ひとつ目のテストデータの少なさに関しては，データを収集して増やす，もしくは学習データと入れ替えて複数回試行する必要がある．ふたつ目に関しては，UCI Machine Learning Repository に公開されているデータセットの内，比較的分類が容易なデータを用いて同様に人工的にコンセプトドリフト用のテストデータとすることが考えられる．みっつ目に関しては，異なるコンセプトドリフト検出法として、Patra ら[Patra 07]の SOM を用いた教師なし変化点検出法の適用を検討する．

4. まとめ

本論文では，コンセプトドリフト適応型学習ベースモニタリングシステムの実現を目指し、ユーザとのインタラクションが必要で低速な学習機能と，オンラインでの高速な処理を実現するデータストリーム管理システム(DSMS)を組み合わせるアーキテクチャを提案した．この手法では，統計に基づいた動的な閾値管理手法によりコンセプトドリフトを検出し、適切なタイミングでモデルの更新を行うことで，持続的なモニタリングを目指す．この手法実現の第一歩として，コンセプトドリフト検出性能の評価を目的とした実験を行った．木材破損時の音波データを SOM のマイクロクラスタへ逐次的に割り当て，その時の割り当て誤差の変化を観察した．なお，この時，データの入力順を作為的に並び替えることで，人工的にコンセプトドリフトを起こした．結果として，コンセプトドリフト前後で，割り当て誤差に優位な差は見られなかった．そこで，今後，学習データと同じ種類のテストデータを増やすもしくは学習データと入れ替えて複数回試行する，対象データを木材破損時のものから変更する，割り当誤差以外のコンセプトドリフト検出基準を検討する，などの変更を加えて同様の実験を行い，コンセプトドリフトの適切な検出を目指す．

その後、DSMS として用いる Odysseus への実装、Labeling プロセスなどの構築を行うことで、提案アーキテクチャの実現を予定している．さらに，提案アーキテクチャを酸化物型燃料電池(SOFC)やその他構造物の物理的損傷に対するモニタリングシステムとして適用する．

謝辞

本研究の一部は「附置研究所間アライアンスによるナノとマクロをつなぐ物質・デバイス・システム創製戦略プロジェクト」特別経費（文部科学省）の助成を受けて行われた．

参考文献

[Zliobaite 10] Indre Zliobaite, Learning under Concept Drift: an Overview, http://arxiv.org/abs/1010.4784, (2010)

[Gama 04] J.Gama, Pedro Medas, Gladys Castillo, Pedro Rodrigues: Learning with Drift Detection, *Proceedings of 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil,* pp. 286-295 (2004)

[Silva 13] Jonathan A. Silva, Elaine R. Faria, Rodrigo C. Barros, Eduardo R. Hruschka, Andre C. P. L. F. DE Carvalho, Data Stream Clustering: A Survey: *ACM Computing Surveys*, Vol. 4, Issue 1, 13, (2013)

[Appelrath 12] H.-Jurgen Appelrath, Dennis Geesen, Marco Grawunder, Timo Michelsen, Daniela Nicklas: Odysseus – A Highly Customizable Framework for Creating Efficient Event Stream Management Systems, *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, pp. 367-368 (2012)

[福井 10] 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)

[北川 11] 北川哲平, 福井健一, 佐藤一永, 水崎純一郎, 沼尾正行:キーグラフと SOM を用いた稀な重要事象抽出による燃料電池の損傷評価, 情報処理学会論文誌:数理モデル化と応用, Vol. 4, No. 2, pp. 1-12 (2011)

[佐藤 05] 佐藤一永, 橋田俊之, 八代圭司, 湯上浩雄, 川田達也, 水崎純一郎:模擬作動環境下における固体酸化物燃料電 池の機械的損傷評価法の開発, *Journal of the Ceramic Society of Japan*, Vol. 113, pp. 562-564 (2005)

[Andras 02] Andras,P., Kernel-Kohonen networks: *International Journal of Neural Systems,* Vol.12, pp.117-135 (2002)

[Boulet 08] Boulet, R., Jouve, B., Rossi, F. and Villa, N.: Batch Kernel SOM and Related Laplacian Method for Social Network Analysis, *Neurocomputing*, Vol. 71, pp. 1257-1273 (2008)

[Parta 07] Swarnajyoti Patra, Susmita Ghosh, Ashish Ghosh: Unsupervised Change Detection in Remote-Sensing Images using Modified Self-Organizing Feature Map Neural Network, *IEEE Proceedings of the International Conference on Computing: Theory and Applications*, (2007)