

D-013

スパースコーディングを用いた熱間薄板圧延の異常検知に関する一検討

A Consideration on Abnormality Detection in Hot Strip Rolling using Sparse Coding

永島 元貴† 菅沼 睦々‡ 亀山 渉†††

Motoki Nagashima Mutsumi Sukanuma Wataru Kameyama

手塚 知幸‡ 今成 宏幸‡ 篠永 裕之‡ 新田 勤子‡

Tomoyuki Tezuka Hiroyuki Imanari Hiroyuki Shinonaga Isoko Nitta

1. はじめに

今日、製造ラインでは品質管理のために様々なセンサを多数設置し、センサから出力される時系列データを用いて異常検知等を行っている。しかし、従来の異常検知では、各センサの値を閾値で判別する手法が主として用いられており、あらかじめ適切な閾値を定めるためには多くのコストや熟練したオペレータの経験が必要である。そこで、本稿ではスパースコーディングを用い、上記の点を改善する異常検知手法を検討した。

スパースコーディングは、式(1)のように入力信号 \mathbf{x} を辞書(基底行列) \mathbf{D} ($m \times n, m \ll n$)の少ない基底の線形和で近似表現する手法である。

$$\mathbf{D}\boldsymbol{\alpha} = \mathbf{x} \quad (1)$$

スパースコーディングでは、スパース係数 $\boldsymbol{\alpha}$ の多くの係数が0となるため、信号を効率良く表現できる。また、あらかじめ規定された基底ではなく、学習した基底を用いることでデータに適した柔軟な表現が可能である。本手法を異常検知に適用した例として文献[1, 2]の報告があり、その有用性が示されている。今回、センサデータ解析への適用を考え、正常時の観測データを学習して基底を構築することで、異常検知するシステムの検討を行った。

2. 実験

2.1 入力信号

本実験に使用したデータは、プラント所有者の許可を得て採取した、実際の熱間薄板圧延の製造ラインで出力された375次元の時系列センサデータである。各センサの観測データは50ミリ秒間隔で出力され、約1,700,000点収集した。これは約1日分のデータに相当する。ここで、異常は圧延過程中的のみ発生するとし、実験には圧延中のみのデータを使用した。なお、これを正常時のデータとし、観測データ \mathbf{x}_n と定義する。観測データ \mathbf{x}_n は約320,000点であった。

次に、実際に起こりうる異常時のデータを人工的に生成した。異常は、圧延中の異常振動を想定し、ある重みを持った正弦波として与え、375次元の内の1つのセンサ出力値のみに発生するものとした。これを観測データに加え、異常時のデータとし、以下、擬似異常データ \mathbf{x}_a と呼ぶ。

2.2 実験手順

図1に実験手順のフローを示す。まず、入力信号として、正常時のデータである観測データ \mathbf{x}_n を与え、辞書 \mathbf{D}_n を学習

† 早稲田大学大学院国際情報通信研究科

†† 早稲田大学国際情報通信研究センター

††† 早稲田大学基幹理工学部情報通信学科

‡ 東芝三菱電機産業システム株式会社 (TMEIC)

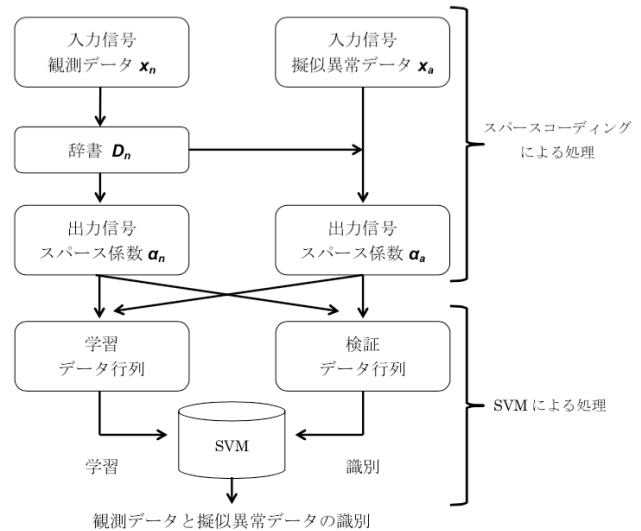


図1 実験のフローチャート

する。次に、辞書 \mathbf{D}_n に対する観測データ \mathbf{x}_n のスパース係数 $\boldsymbol{\alpha}_n$ と擬似異常データ \mathbf{x}_a のスパース係数 $\boldsymbol{\alpha}_a$ を求める。以下、観測データ \mathbf{x}_n から得られたスパース係数 $\boldsymbol{\alpha}_n$ を「正常時のスパース係数」と呼び、擬似異常データ \mathbf{x}_a から得られたスパース係数 $\boldsymbol{\alpha}_a$ を「異常時のスパース係数」と呼ぶ。

最後に、正常時のスパース係数 $\boldsymbol{\alpha}_n$ を正常とし、異常時のスパース係数 $\boldsymbol{\alpha}_a$ を異常として、これらのスパース係数をSVM (Support Vector Machine) によって学習させる。ここで、正常時のスパース係数 $\boldsymbol{\alpha}_n$ と異常時のスパース係数 $\boldsymbol{\alpha}_a$ から、それぞれランダムサンプリングによってデータ全体の90%をSVMの学習用データとして、残り10%を実験用の識別データとして用いた。

2.3 評価方法

提案手法を評価するため、SVM単独の結果と比較を行う。具体的には、異常のない観測データ \mathbf{x}_n を正常とし、擬似異常データ \mathbf{x}_a を異常とし、提案手法のSVM識別と同様、それぞれ90%を学習用データとして用い、残りの10%を実験用の識別データとして用いた。このSVM単独による識別結果と提案手法で得られた識別結果の比較を行い、評価する。

2.4 実験環境

本実験において、スパースコーディングの実装には、RパッケージSPAMS[3, 4]を用いた。またSVMの実装は、Rのkernlabパッケージ[5]を用いた。カーネルはラジアル基底関数カーネルを使用した。

スパースコーディングの制約上、辞書は入力信号の次元数375よりも大きい過完備基底となるため、基底数は約2.5倍の1000とし、辞書構築および信号分解にかかる正規化パラメータ λ の値は0.15に設定した。

2.5 実験結果

提案手法の実験結果の一例と10交差検定の平均識別率を表1に、SVM単独によるそれぞれの結果を表2に示す。

表1 提案手法による結果

		識別結果	
		正常	異常
入力信号	正常	31457	520
	異常	0	21097
識別率		0.99020	
10交差検定の平均識別率		0.98994	

表2 SVM単独による結果

		識別結果	
		正常	異常
入力信号	正常	31366	599
	異常	0	21109
識別率		0.98871	
10交差検定の平均識別率		0.98911	

SVM単独による10交差検定の平均識別率は約98.91%であり、提案手法による10交差検定の平均識別率は約98.99%であった。また、両手法とも異常であるものを正常として識別するものはなかったが、正常なものを異常と誤判別したものが、SVM単独で行うよりも提案手法では平均約8%減少し、識別精度が向上していることがわかった。

スパースコーディングの結果から、得られるスパース係数 α_n 、 α_a の全体のスパース性を解析した。ここでスパース性とは、スパース係数 α_n 、 α_a の全体に占める0の割合である。辞書 D_n は、擬似異常データ \mathbf{x}_a のパターンを学習していないことから、正常時に比べて異常時のスパース係数 α_a のスパース性は減少すると予測したが、正常時のスパース係数 α_n のスパース性は91.972%であり、それに対して、異常時のスパース係数 α_a のスパース性は91.971%となり、両者に明確な差異はなかった。

そこで、個々のスパース係数中の0の個数を計算し、その個数の0を持つスパース係数の数をヒストグラムで分析した。その結果、正常時と異常時、それぞれのスパース係数 α_n 、 α_a が持つ0の分布がやや異なることがわかった(図2参照)。

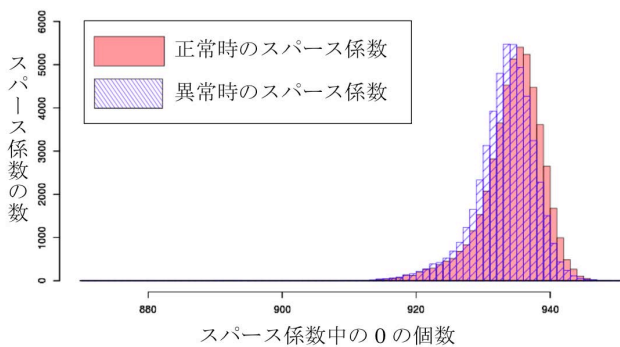


図2 スパース係数中の0の分布

3. 考察

実験結果から、疑似異常データに関しては、提案手法とSVM単独の両者とも、正しく異常として識別された。正常を異常と誤判別してしまったものに関しては、スパース係数をSVMで識別する提案手法の方が減少した。要因としては、スパースコーディングによって観測データの特徴を含む重みがスパース係数に現れ、SVMの結果に作用したと考えられる。

また、使用した辞書 D_n を用いて異常時と正常時の係数ベクトルを観測データに復元したが、両者とも元の観測データとの大きな差異は見られなかった。このことは、スパー

コーディングが観測データを正確に学習したことを意味している。

正常時のスパース係数 α_n と異常時のスパース係数 α_a に関してスパース性を分析した結果からは、スパース係数のスパース性にほとんど違いはなかった。これは、用意した疑似異常データが正常データに対して大きな差異がなかったためと推定される。しかしながら、データの再現性は高いことから、ある基底の組み合わせによって元データを精度よく表現していると考えられる。その結果、図2に示すように、正常時と異常時でスパース係数1データあたりの0の個数分布が異なると推測する。

4. まとめと今後の課題

スパースコーディングによって観測データ \mathbf{x}_n を学習する辞書 D_n を構築し、観測データ \mathbf{x}_n から算出した正常時のスパース係数 α_n と疑似異常データ \mathbf{x}_a から算出した異常時のスパース係数 α_a を用いて異常検知するための手法を提案し、SVM単独による結果と比較し評価を行った。実験の結果、提案手法では、SVM単独より識別精度が向上することが確認できた。今回は、教師あり機械学習であるSVMを併用したが、スパースコーディングを適用することにより、従来の異常検知で用いられる閾値の設定に依存しないシステムを構築できる可能性を示唆した。

しかし、SVMを加えることによって識別精度が向上した理由についてはまだ明確ではない。そのため、スパースコーディングで学習した基底の解析が必要である。また、今回実験に用いた異常データは、人工的に生成した疑似異常であるため、現実に対応できるかは未知である。今回使用したものは異なる実際に起こりうる異常でも、本手法で同様の結果が得られるか検証が必要である。

参考文献一覧

- [1] B. Zhao, F.F. Li and E.P. Xing, "Online Detection of Unusual Events in Videos via Dynamic Sparse Coding", IEEE Conference on Computer Vision and Pattern Recognition, pp. 3313-3320, 2011
- [2] A. Adler, M. Elad, Y. Hel-Or and E. Rivlin, "Sparse coding with abnormal detection", IEEE International Workshop on Machine Learning for Signal Processing, pp. 1-6, 2013
- [3] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding", Journal of Machine Learning Research, Vol. 11, pp. 19-60, 2010
- [4] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online Dictionary Learning for Sparse Coding", International Conference on Machine Learning, Montreal, Canada, 2009
- [5] <http://cran.r-project.org/web/packages/kernlab/kernlab.pdf> (2015年6月10日 最終確認)