

# 構造正則化学習を用いた混雑シーンにおける異常検知

Abnormal Event Detection in Crowded Scenes using Structured Sparse Learning

掃部健\*<sup>1</sup>      河原吉伸\*<sup>1</sup>      鷲尾隆\*<sup>1</sup>  
Takeru Kamon      Yoshinobu Kawahara      Takeshi Washio

\*<sup>1</sup>大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

In recent years, abnormal event detection in crowded scenes has been widely studied in computer vision. In this paper, we propose a novel abnormality detection algorithm for crowded scenes in video using structured sparse coding, where we incorporate the spatio-temporal proximity in video into detection with generalized fused regularization. We show some empirical examples, where the proposed algorithm works better than the state-of-the-art algorithm.

## 1. はじめに

近年、多くの人や物が混み合った状況を映した混雑シーンにおける異常検知が盛んに研究されている。また、これらのデータの中には、学習に利用可能な離散構造が存在する 경우가多く、代表的なものとしては、変数間のグラフ構造やグループ構造、または階層的構造などが挙げられる。本研究では、混雑シーンの時空間的な画素の近接性に着目し、構造を正則化学習の中で明示的に利用するための構造正則化学習と呼ばれる枠組みを利用する。本稿では、構造正則化学習を用いた混雑シーンにおける異常検知を対象とし、従来よりも異常検知性能が高く、高速な計算を可能とする手法を提案する。

我々と同じく、構造正則化学習を用いた混雑シーンの異常検知を扱った研究として、Congら [1] があり、Congらは Group LASSO [2] を利用することで、各画素についての時系列データをグループとした構造を組み込んだ正則化学習を提案した。これにより、正常なデータを表すモデルの生成に必要な特徴を画素単位で選ぶことができ、動画の時間方向の構造を組み込んだ学習が可能となる。本研究では、さらに動画の空間方向で隣接した画素の関係に着目し、動画の時間・空間方向の構造を正則化学習に組み込む手法を提案する。

本稿の構成は以下である。2. で、提案手法について述べ、3. で評価実験について述べ、4. で本稿のまとめとする。

## 2. 提案手法

### 2.1 訓練データ

まず、提案手法で用いる訓練データについて述べる。本研究は動画の異常検知を対象にするため、画像の時系列データを訓練データとして用いる。

時間の経過とともに観測された画像の画素に対応する変数を  $x_{ij}(t)$  ( $i = 1, \dots, p, j = 1, \dots, q, t = 1, \dots, T$ ) と定義する。ここで  $p, q$  はそれぞれ画像の縦、横の画素位置に対応する。ある時点  $t$  で画像のある行  $i$  に関して観測された各変数  $x_{ij}(t)$  を  $q$  個並べた行ベクトルを  $x_i(t)$  とし、さらに、ある時点  $t$  における行  $i = 1, \dots, p$  の行ベクトル  $x_i(t)$  を束ねた  $1 \times pq$  の行ベクトルを  $X(t) = [x_1(t) \dots x_p(t)]$  とする。 $X(t)$  は時点  $t$  における画像 1 枚のデータである。

次に  $\tau$  ( $\tau > 1$ ) 枚の画像データをまとめた行ベクトル  $Y(t) = [X(t) \dots X(t-\tau+1)]$  を定義すると、提案手法で用いる訓練データは  $T-\tau+1$  個の行ベクトル  $Y(t)$  を時間順序に束ねた  $(T-\tau+1) \times d$  行列  $Y$  で表される (ただし  $d = pq\tau$ )。

### 2.2 Generalized Fused LASSO

提案手法では、動画の時間・空間方向の構造を正則化学習に組み込むために、Generalized Fused LASSO [3] を利用する。学習に組み込む構造として、係数の各要素がグラフのノードに対応するようなノード  $V$ 、係数の各要素間の関係に対応するエッジ  $E$  のグラフ  $G = (V, E)$  を考えたとき、係数ベクトル  $w = (w_1, \dots, w_M)^T$  とすると、Generalized Fused LASSO の解は式 (1) により与えられる。

$$\min_w l(w) + \lambda_1 \sum_{m=1}^M |w_m| + \lambda_2 \sum_{(i,j) \in E} |w_i - w_j| \quad (1)$$

式 (1) の第二項により、係数の要素のスパース性が高まり、第三項により、隣接するノードに対応する係数の要素の推定値が互いに近い値になりやすくなるため、この手法では、隣接するノードに対応する要素の推定値の差異が 0 になりやすくなる。

### 2.3 定式化

訓練データ  $Y \in \mathbb{R}^{(T-\tau+1) \times d}$  と、グラフ  $G = (V, E)$  が与えられたとき、係数行列  $C \in \mathbb{R}^{d \times d}$  とすると、提案手法は Generalized Fused LASSO の拡張である以下の式 (2) で定式化される。

$$\min_C \|Y - YC\|_2^2 + \lambda_1 \sum_{i,j} |C_{ij}| + \lambda_2 \sum_{(ij, i'j') \in E} |C_{ij} - C_{i'j'}| \quad (2)$$

ここで、図 1 は時系列データ  $Y(t)$  と、ある画素に対するグラフ内の構造を表す。赤い点はノードを、各ノードをつなぐ線はエッジを示し、黒いエッジは動画の時間方向の構造、青いエッジは動画の空間方向の構造に対応する。既存手法は黒いエッジが示す動画の時間方向の構造のみ組み込んだ学習であり、提案手法ではさらに青いエッジが示す動画の空間方向の構造を組み込んだ学習を行う。この青いエッジにより、周辺の画素のデータの影響も考慮した学習を行えるため、動画の空間方向の人や物の動きを捉えることができるようになる。この構造を式 (2) の第三項に組み込むことにより、隣接するノードに対応する要素の推定値の差異を 0 に縮小推定することができる。

連絡先: 掃部健, 大阪大学産業科学研究所, 567-0047 大阪府茨木市美穂が丘 8-1-N502-2, kamon@ar.sanken.osaka.ac.jp

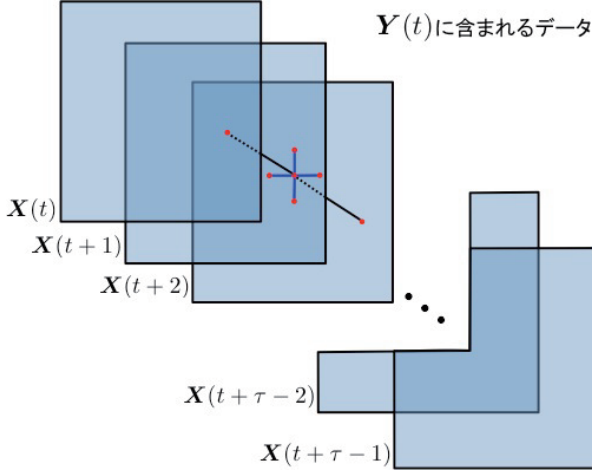


図 1: 時系列データ  $Y(t)$  におけるグラフ  $G = (V, E)$

## 2.4 最適化手法

定式化された問題において、係数行列  $C$  の次元は  $d \times d$  であった。実際の混雑シーンの異常検知に用いられるデータは少なくとも縦、横 100 画素程度の大きさが必要になるため、 $\tau$  を必要最低限の  $\tau = 2$  に設定しても係数行列  $C$  の次元は 2 万  $\times$  2 万と大きく、計算に時間がかかる。そのため、本研究では、係数行列  $C$  を  $C = \alpha\beta^T$  のようにすることで次元を落とし、計算時間を短縮する。ただし  $\alpha \in \mathbb{R}^{d \times r}$ ,  $\beta \in \mathbb{R}^{d \times r}$  ( $r$  は正の整数) とする。この置換は係数行列  $C$  が低ランクであるため、十分に良い近似を得られる。

これにより、式 (2) で定式化された問題は式 (3) で表される。

$$\min_{\alpha, \beta^T} \|Y - Y\alpha\beta^T\|_2^2 + \lambda_1 \sum_{i,j} |\alpha_{ij}\beta_{ij}| + \lambda_2 \sum_{(ij, i'j') \in E} |\alpha_{ij}\beta_{ij}^T - \alpha_{i'j'}\beta_{i'j'}^T| \quad (3)$$

式 (3) の最適解は、まず  $\beta$  を固定し  $\alpha$  について最適化、次に  $\alpha$  を固定し  $\beta$  について最適化という計算を繰り返すことで求める事ができる [1]

まず  $\alpha$  についての最適化について述べる。 $\beta$  を固定すると、式 (3) は式 (4) となる。

$$\min_{\alpha} \|Y - Y\alpha\beta^T\|_2^2 + \lambda_1 \sum_{i,j} |\alpha_{ij}| + \lambda_2 \sum_{(ij, i'j') \in E} |\alpha_{ij} - \alpha_{i'j'}| \quad (4)$$

式 (4) の最適化は、パラメトリックフローに帰着することができ、最悪の計算量  $O(dr|E|\log(d^2r^2/|E|))$  での高速な計算が可能である [4]。

次に  $\beta$  についての最適化について述べる。 $\alpha$  を固定すると、式 (3) は式 (5) となる。

$$\min_{\beta} \|Y - Y\alpha\beta^T\|_2^2 \quad (5)$$

この最適解は  $\beta$  の各列  $\beta_{\cdot j} (j = 1, \dots, r)$  について  $j = 1$  から  $j = r$  まで順番に式 (6) で更新することで求めることができる。 $u$  は  $d \times 1$  の列ベクトルである。

$$\beta_{\cdot j} = \text{sgn}(u) \odot \min(u, 1) \quad (6)$$

$$u = \frac{(Y - \sum_{j' \neq j} (Y\alpha)_{\cdot j'} \beta_{\cdot j'}^T)^T (Y\alpha)_{\cdot j}}{\|(Y\alpha)_{\cdot j}\|^2}$$

ここで、 $\odot$  は  $(a \odot b)_i = a_i b_i$  で定義される演算子である。 $\text{sgn}(u)$  は  $u$  の要素を  $u_i (i = 1, \dots, d)$ ,  $a = \text{sgn}(u)$  としたとき、それぞれの  $i$  に対して  $u_i > 0$  のとき  $a_i = 1$ ,  $u_i = 0$  のとき  $a_i = 0$ ,  $u_i < 0$  のとき  $a_i = -1$  を返す関数であり、 $a$  は  $d \times 1$  の列ベクトルとなる。 $\min(u, 1)$  は  $b = \min(u)$  としたとき、それぞれの  $i$  に対して  $u_i \geq 1$  のとき  $b_i = 1$ ,  $u_i < 1$  のとき  $b_i = u_i$  を返す関数であり、 $b$  は  $d \times 1$  の列ベクトルとなる。

提案手法の疑似コードを Algorithm1 に示す。

### Algorithm 1 提案手法の疑似コード

**Input :**  $Y \in \mathbb{R}^{(T-\tau+1) \times d}$ ,  $\alpha_0 \in \mathbb{R}^{d \times r}$ ,  $\beta_0 \in \mathbb{R}^{d \times r}$ ,  $\lambda_1 > 0$ ,  $\lambda_2 > 0$

**Output :**  $(\alpha\beta)^*$

**Step.k** ( $k \geq 0$ )

$\beta_k$  を固定し、式 (7) をパラメトリックフロー問題に帰着し、 $\alpha$  について最適化する。

$$\alpha_{k+1} = \min_{\alpha} \|Y - Y\alpha\beta_k^T\| + \lambda_1 \sum_{i,j} |\alpha_{ij}| + \lambda_2 \sum_{(ij, i'j') \in E} |\alpha_{ij} - \alpha_{i'j'}| \quad (7)$$

次に  $\alpha_{k+1}$  を固定し、 $\beta$  の各列  $(\beta_{k+1})_{\cdot j} (j = 1, \dots, r)$  について  $j = 1$  から  $j = r$  まで順番に式 (8) で更新する。

$$(\beta_{k+1})_{\cdot j} = \text{sgn}(u) \odot \min(u, 1) \quad (8)$$

$$u = \frac{(Y - \sum_{j'=1}^{i-1} (Y\alpha_{k+1})_{\cdot j'} (\beta_{k+1})_{\cdot j'}^T)^T (Y\alpha_{k+1})_{\cdot j}}{\|(Y\alpha_{k+1})_{\cdot j}\|^2} - \frac{\sum_{j'=i+1}^r (Y\alpha_{k+1})_{\cdot j'} (\beta_{k+1})_{\cdot j'}^T (Y\alpha_{k+1})_{\cdot j}}{\|(Y\alpha_{k+1})_{\cdot j}\|^2}$$

## 3. 評価実験

### 3.1 実験内容

既存手法、提案手法をそれぞれ計算機で実装し、混雑シーンの実データ (UCSD Ped1 dataset [5][6]) を元に異常検知性能、計算速度を計測し、比較する。今回の実験では混雑シーンに現れる車、自転車、スケートボードなどの乗り物を異常として検知することを目的とする。そのため訓練データには乗り物が現れない正常なデータを用い、そこから得たパラメータを用いてテストデータに現れる乗り物を検知する。異常検知性能は、訓練データ  $Y$  と同じ形式のテストデータ  $Z$  に対する二乗和誤差  $\|Z - ZC\|_2^2$  によって評価した。

### 3.2 実験結果

表 1 は既存手法、提案手法それぞれの計算時間を示した表である。

図 2 はテストデータに対する既存手法の二乗和誤差を縦軸に、時間  $t$  を横軸にとったグラフであり、図 3 はテストデータに対する提案手法の二乗和誤差を縦軸に、時間  $t$  を横軸にとったグラフである。図 4 は図 2、図 3 のグラフに対する ROC 曲線を示したものである。ここで、図 4 の ROC 曲線の説明をす

る。まず、図2、図3の縦軸である二乗和誤差に対して閾値を設定し、二乗和誤差が閾値より大きい場合は異常(陽性)、閾値よりも小さい場合は正常(陰性)と予測する。今回の実験ではテストデータ Test001 において乗り物が現れる  $t = 80$  から  $t = 150$  の間を真の陽性、その他の  $t$  では真の陰性としておく。ROC 曲線は、横軸を False Positive Rate(真の陰性を陽性と予測する割合)、縦軸を True Positive Rate(真の陽性を陽性と予測する割合)とし、閾値を変化させてプロットしたものである。図4のROC曲線が(0,1)の近くを通るほど異常検知性能が高いことを意味する。

### 3.3 考察

図4を見ると、提案手法は既存研究と同等、もしくはそれ以上の異常検知性能があることがわかる。この点については、提案手法は既存研究に動画の空間方向の構造を加えたものであるので、妥当な結果だと思われる。

また、二乗和誤差の時系列データを示す図2、図3は、ほぼ同じグラフとなっているが、 $t = 100, 140$  付近の二乗和誤差と  $t = 130$  付近の二乗和誤差を比較すると、提案手法の方がその差が小さく、 $t = 130$  付近でも  $t = 100, 140$  付近と同等の異常検知性能があることがわかる。

図6を見ると、異常として検知すべき自転車が歩行者と重なって映っている。これは既存研究では検知が難しく、提案手法では検知が可能なケースであるため、提案手法のグラフでは  $t = 130$  付近でも高い異常検知性能を示していると考えられる。これらの結果から、提案手法は混雑シーンにおいて既存研究よりも高い異常検知性能があることがわかった。

計算速度については表1から明らかであり、提案手法では  $C = \alpha\beta^T$  の置換により次元を落とし、高速なパラメトリックフロー問題に帰着して計算することで既存研究に比べて大幅な計算時間の短縮を実現することができた。

## 4. まとめ

本研究では構造正則化化学習を用いた混雑シーンの異常検知に関する研究を行い、評価実験によって、異常検知性能、計算速度ともに既存手法より優れた結果を得ることができた。異常検知は応用範囲が広く、様々な分野で利用されるものであるため、異常検知性能、計算速度を上げることが出来れば、それらの分野の発展にも密接に関与してくると思われる。今後の課題としては、正則化パラメータ  $\lambda$  の最適化、異常な場所の同定、オンラインでの学習などが挙げられる。

実験	計算時間(秒)
既存手法	202368
提案手法	8989

表1: 実験1, 実験2の計算時間

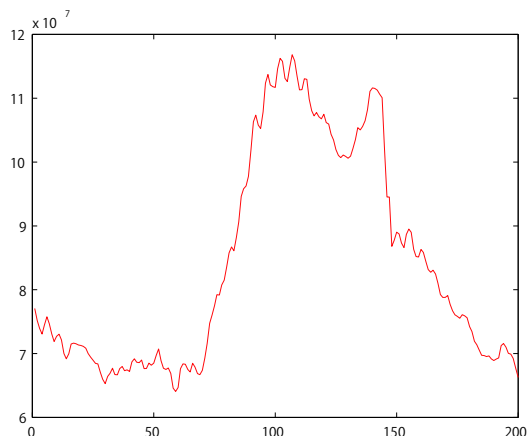


図2: 横軸:t, 縦軸:既存手法の二乗和誤差

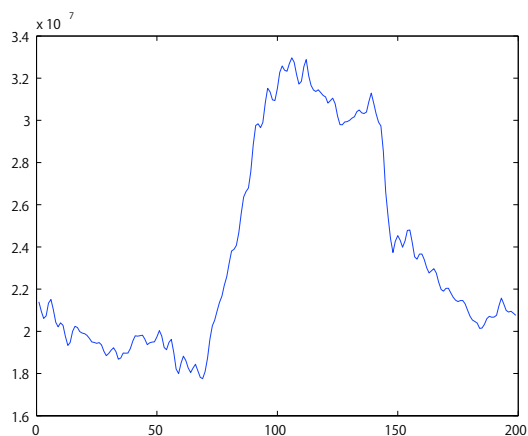


図3: 横軸:t, 縦軸:提案手法の二乗和誤差

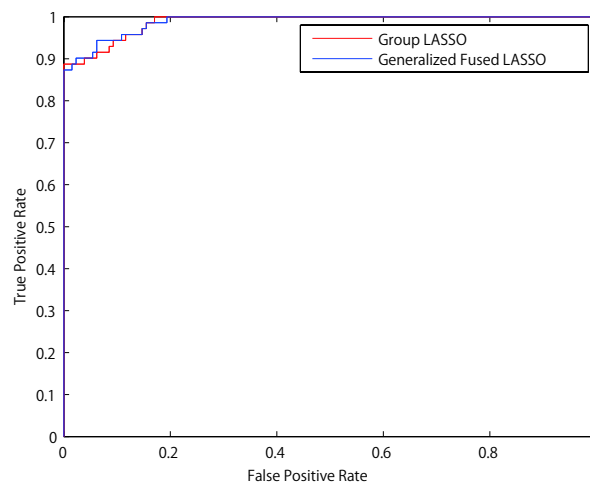


図4: 図2, 図3に対するROC曲線



図 5: テストデータの画像 ( $t = 100$ )



図 6: テストデータの画像 ( $t = 130$ )



図 7: テストデータの画像 ( $t = 140$ )

## 参考文献

- [1] Y. Cong, J. Yuan, J. Liu. Sparse reconstruction cost for abnormal event detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 2449-3456.
- [2] M. Yuan, and Y. Lin. Model selection and estimation in regression with grouped variables. J. Roy. Statist. Soc. Ser. B, 68, 49-67, 2006.
- [3] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pp. 267-288, 1996.
- [4] G. Gallo, et al. A fast parametric maximum flow algorithm and applications. SIAM Journal of Computing 18(1): 30-55, 1989.
- [5] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In CVPR, 2010, pp. 1975-1981.
- [6] (<http://www.svcl.uscd.edu/projects/anomaly>)