

平成24年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	青 木 栄 太
卒業研究題目	利用可能な教師に応じた 適応的線形次元圧縮の提案	

近年、コンピュータ、スマートフォン等の機器が一般家庭に普及し、一般ユーザが画像等のマルチメディアデータに接する機会が増えてきた。しかし画像等の高次元データは、高い冗長性を持っており、また高次元に由来する「次元の呪い」問題も存在する。このため、元データの本質的な構造を保ちつつ、次元数を削減し、その後の処理を効率化する次元削減手法に対する要求は高い。一方で、インターネットや電子機器の性能向上に伴い、処理対象となるデータ量は飛躍的に増加している。しかし、その大多数は人手による精査を経ていない、ラベルなしデータとなっており、多数のラベルなしデータとごく少数のラベル付きデータを対象として処理を行うことがしばしばである。

そこで本研究では、与えられた情報を最大限に利用した次元削減を目的として、教師なし次元圧縮手法と部分教師付き次元圧縮手法を、与えられた教師に応じて適応的に統合する手法 ILDR を提案した。具体的には、式 1,2 のように重み α, β を用いて統合する。

Σ_t は全データの分散・共分散行列、 \mathbf{I} は単位行列、 $\Sigma_{b_supervised}$ は教師のみの局所的な分散行列、 $\Sigma_{b_estimated}$ 、 $\Sigma_{w_estimated}$ は教師から全データのカテゴリを予測して計算した級間分散行列、級内分散行列である。

$$\Sigma_{b_integrated} = \beta \times \Sigma_t + (1 - \beta) \times (\alpha \times \Sigma_{b_supervised} + (1 - \alpha) \times \Sigma_{b_estimated}) \quad (1)$$

$$\Sigma_{w_integrated} = \beta \times \mathbf{I} + (1 - \beta) \times \Sigma_{w_estimated} \quad (2)$$

実験では、実画像から得られる特徴データと多峰性人工データ、それぞれに対して ILDR のパラメータを変えて比較を行った。評価尺度は、圧縮後のカテゴリ間の分離度を示す判別評価基準と、最近傍法による識別率の2通りで行った。実データに関しては、図1より、教師数が少ない時は β が高く、教師数が多い時は、 β が低いほど判別評価基準が高いことがわかった。図2では、最近傍法での識別に対しては ILDR の有効性が確認できなかった。多峰性データに関しては、図3より、LFDA の有効性は確認できたが、ILDR の有効性は確認できなかった。このことから、今後は ILDR に対する適切な評価方法とサンプルデータを検討する必要があると考えられる。

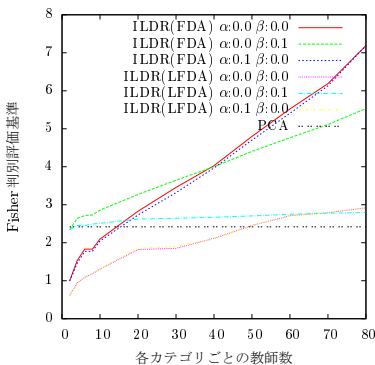


図1: 実データに対する判別評価基準

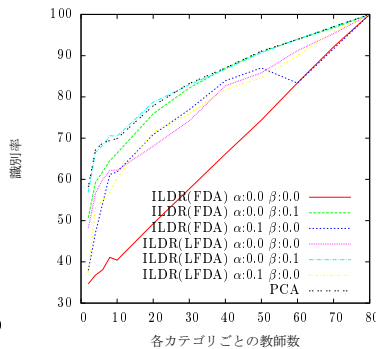


図2: 実データに対する識別率

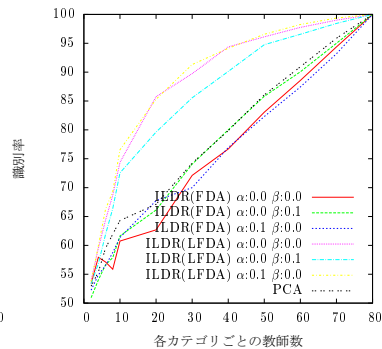


図3: 多峰性人工データに対する識別率