1R-09

パラメータ共有型マルチモーダル深層自己符号化器を用いた部分観測下多様体学習の検討

佐々木　勇人†　　中田　雅也‡　　濱津　文哉‡　　濱上　知樹‡
†横浜国立大学大学院工学府　　‡横浜国立大学大学院工学研究院

1　はじめに

実世界のシステムを制御するにあたって，観測データから制御対象の状態を自律的に推定することが求められる。本研究では，高次元データに対する特徴抽出が可能である深層自己符号化器を特徴抽出器として用い，その特徴をシステムの状態として扱う。

しかしながら，観測データから抽出された特徴を状態として扱う際にはその観測が部分観測であることに起因して，一つの観測が複数の状態にエイリアスしてしまう。

本研究では，相補的な複数種類の観測データ (マルチモーダルデータ) を用いることで，エイリアスの影響を緩和する。すでにマルチモーダルデータに対して深層自己符号化器を利用する研究 [1] があるが，この研究ではモダリティ間で同期のとれたマルチモーダルデータを利用している。一方，本研究では非同期マルチモーダルデータから特徴を抽出する。これは，現実には同期を取ることが難しいと考えたためである。同期は取れていないものの，訓練時にモダリティの異なる観測データを補助的に用いることで，単一モダリティでは捉えることが難しかった特徴を学習することを目指す。

2　提案手法

本研究では，各モダリティに対応する深層自己符号化器間でネットワークの学習パラメータを相互参照しながら再構築誤差を最小化する。その相互参照の方法として本研究では Shu[2] らの研究で提案された Weakly-shared な手法を用いる。この手法では，学習パラメータがモダリティ間で互いに近づくようにパラメータ間の差のフロベニウスノルムを最小化する。“Weakly” とあるように，パラメータを完全に一致させるのではなく，ある程度のずれは許容してパラメータを共有する。

深層自己符号化器は，層ごとに写像を繰り返しながら入力データを再構築するが，これは観測空間に埋め込まれた多様体を繰り返し変形および射影していることになる。各層の学習パラメータは写像を表現しているため，モダリティ間でパラメータを共有するということは写像を共有することにほかならない。深層自己符号化器はエンコーダとデコーダからなるが，エンコーダ上位層のパラメータを共有することは，特徴空間に対応する多様体の形状に大きな影響を与える。

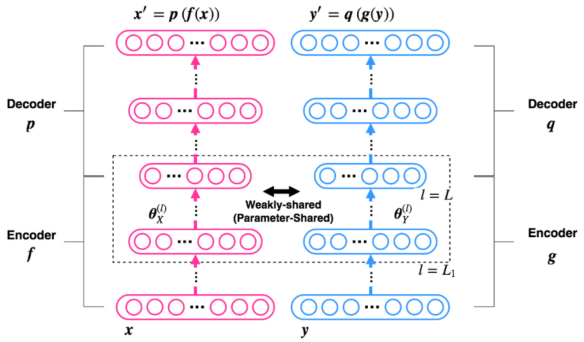

図 1: Weakly-shared Deep Autoencoders. Parameters (such as weights and bias) are shared among networks.

図 1 のネットワークモデルは 2 つのモダリティ X および Y に対応する深層自己符号化器に Weakly-shared な手法を適用することで各モダリティの特徴空間を学習することを目指しており，次式 (1)，(2) の目的関数により最適化する。

$$\min_{\Theta} E_x\left[\mathcal{L}\left(\boldsymbol{x}, \boldsymbol{p}\left(\boldsymbol{f}(\boldsymbol{x})\right)\right)\right] + \frac{\lambda}{2}\sum_{l=L_1+1}^{L}\sum_{\boldsymbol{\theta}_X^{(l)},\boldsymbol{\theta}_Y^{(l)}} \|\boldsymbol{\theta}_X^{(l)} - \boldsymbol{\theta}_Y^{(l)}\|_F^2 \quad (1)$$

$$\min_{\Theta} E_y\left[\mathcal{L}\left(\boldsymbol{y}, \boldsymbol{q}\left(\boldsymbol{g}(\boldsymbol{y})\right)\right)\right] + \frac{\lambda}{2}\sum_{l=L_1+1}^{L}\sum_{\boldsymbol{\theta}_X^{(l)},\boldsymbol{\theta}_Y^{(l)}} \|\boldsymbol{\theta}_X^{(l)} - \boldsymbol{\theta}_Y^{(l)}\|_F^2 \quad (2)$$

$\boldsymbol{\theta}_X^{(l)}, \boldsymbol{\theta}_Y^{(l)}$ はモダリティ X と Y それぞれのネットワークにおける l 層目の学習パラメータであり，L_1+1 層から L 層 (エンコーダの最上位層) のパラメータを共有する。また，$\mathcal{L}$ は再構築誤差を計算する関数であり，$\boldsymbol{f}$ と $\boldsymbol{p}$ がモダリティ X に関する深層自己符号化器のエンコーダ及びデコーダ ($\boldsymbol{g}$，$\boldsymbol{q}$ に関しても同様)，$\lambda(\geq 0)$ が学習パラメータの共有度合いを調整するためのハイパーパラメータである。なお，提案手法の目的関数は各モダリティごとに再構築誤差を最小化しており，非同期のマルチモーダルデータに対しても利用可能である。

Investigation for manifold learning under partial observation using parameter shared multimodal deep autoencoders
†Hayato SASAKI, ‡Masaya TANAKA, ‡Fumiya HAMATSU and ‡Tomoki HAMAGAMI
†Graduate School of Engineering, Yokohama National University
‡Faculty of Engineering, Yokohama National University

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

表 1: Hyper parameters for Deep Autoencoders.

learning rate	1.0×10^{-3}
mini-batch size	500
epochs	200
activation function	Sigmoid
loss function $\mathcal{L}$	Cross Entropy
optimization method	Adam

3 実験

パラメータ共有が学習される特徴空間に与える影響を調べるために次の実験を行った。

実験内容 手書き数字データセット MNIST (訓練データ数:55,000，テストデータ数:10,000) から人工的に作成したマルチモーダルデータから各モダリティの特徴を学習し，その特徴を用いて SVM で数字の分類を行う。

深層自己符号化器により得られた特徴空間を評価するために，SVM のハイパーパラメータ C および RBF カーネルのハイパーパラメータ $\frac{1}{2\sigma^2}$ を変化させた時の分類精度を，パラメータを共有しない場合 ($\lambda = 0$) とパラメータを共有した場合 ($\lambda = 1 \times 10^{-3}$) とで比較した。

データセット 観測空間は $[0, 1]^{784}$ であり，各ピクセルの輝度値を 50%の確率で 1 に置き換えるようなノイズを以下の 2 つの方法で与えた。

- モダリティ X: 画像の左半分にノイズ
- モダリティ Y: 画像の左半分にノイズ

ネットワーク 深層自己符号化器の層ごとのニューロン数は 784 → 600 → 600 → 300 → 150 → 75 → 10 → 75 → 150 → 300 → 600 → 600 → 784 であり，各層には Batch Normalization 層が組み込まれている。それ以外のハイパーパラメータは表 1 のとおりである。なお，パラメータ共有はエンコーダの全ての層に対して行い，重み行列 $\boldsymbol{W}^{(l)}$ および Batch Normalization 層の学習パラメータ $\boldsymbol{\beta}^{(l)}$，$\boldsymbol{\gamma}^{(l)}$ を共有している。

実験結果 学習した特徴空間を可視化した結果が図 2 である。特徴空間は 10 次元であるため，t-SNE により 2 次元平面に写像している (*perplexity* = 40.0)。図 2 ではエイリアスの影響を確認するために'3' と'8' のデータのみを表示した。しかしながら，この可視化結果からはパラメータ共有による明確な差は確認できなかった。

そこで定量的評価として，SVM による数字の分類結果を表 2 に示す。表内の太字で示した部分で，$\lambda = 1 \times 10^{-3}$ の正解率が $\lambda = 0$ の正解率を上回っている。この結果から，パラメータを共有した結果得られた特徴空間は，ソフトマージン (C が小) かつ識別面が複雑ではない ($\frac{1}{2\sigma^2}$ が小) SVM に対して比較的有効に働くことが分かる。

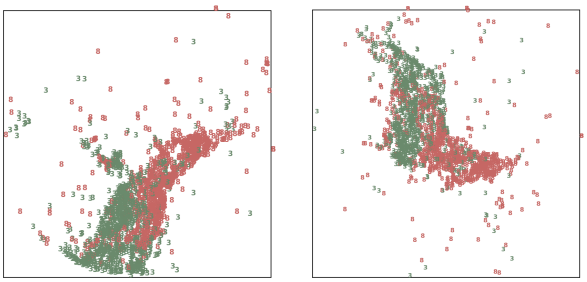

図 2: Feature space of modality X. Only '3' and '8' are visualized (for clarity of aliasing). Left:$\lambda = 0$, Right:$\lambda = 1 \times 10^{-3}$

表 2: Classification accuracies of modality X for SVMs with different hyper parameters.

C \ $\frac{1}{2\sigma^2}$	λ	0.1	1	10
0.1	0	77.40%	87.95%	93.19%
	1×10^{-3}	**84.42%**	**90.07%**	76.01%
1	0	86.84%	91.69%	94.65%
	1×10^{-3}	**87.80%**	**92.60%**	91.49%
10	0	88.78%	93.56%	94.32%
	1×10^{-3}	**90.97%**	92.75%	91.25%

4 おわりに

部分観測下で得られた観測データは，エイリアスにより異なるクラスのデータが混在してしまうため，誤分類せずに識別面を引くことが難しくなってしまう。今回の実験結果は，パラメータ共有によってソフトマージン SVM に対して分類精度が確保できるような特徴空間が学習できたことを示しており，部分観測下でも有効に働くことが示唆された。

参考文献

[1] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 689–696, 2011.

[2] Xiangbo Shu, Guo-Jun Qi, Jinhui Tang, and Jingdong Wang. Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pp. 35–44. ACM, 2015.

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.