

# 深層一般化正準相関分析の提案と 高精度クロスモーダル検索への応用

岩瀬智亮<sup>†</sup> 中山英樹<sup>†</sup>

<sup>†</sup> 東京大学大学院情報理工学系研究科

## 1 はじめに

正準相関分析 (CCA) は、2つの多次元連続値確率変数について線形射影を行うことで、高い相関を持つ線型部分空間の対を見出す手法である。確率変数が3つ以上存在するとき、射影した全ての確率変数対の相関を同時に最大化する線形射影を求める手法は、一般化正準相関分析 (GCCA) と呼ばれる。一般に、マルチタスク学習の観点において各タスクが類似している時には、対応する各学習器が共通要因を獲得することで性能が上がることを期待される。これは、GCCAにおいて入力の確率変数の数を増やすことにより、それぞれの相関がより大きくなるような、各確率変数対の射影が求まることに対応する。本研究では、深層学習のモデルを用いて CCA を非線形拡張した、深層正準相関分析 (DCCA) について、GCCA の最大化基準を取り入れた深層一般化正準相関分析 (Deep Generalized CCA, DGCCA) を提案する。

## 2 正準相関分析の関連手法

### 2.1 正準相関分析 (CCA) [4]

$X = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{p \times N}$ ,  $Y = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{q \times N}$  とする。 $X, Y$  それぞれの分散共分散行列を  $\Sigma_{11}, \Sigma_{22}$ , 交差共分散行列を  $\Sigma_{12}$  とする。また、 $X, Y$  は中心化 ( $\bar{\mathbf{x}} = \bar{\mathbf{y}} = \mathbf{0}$ ) されているものとする。ここで、 $\boldsymbol{\eta} = \mathbf{a}'X \in \mathbb{R}^N$ ,  $\boldsymbol{\phi} = \mathbf{b}'Y \in \mathbb{R}^N$  の線形射影を考えると、射影後の2変数  $\boldsymbol{\eta}, \boldsymbol{\phi}$  の相関係数は (1) で表せる。

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}'\Sigma_{12}\mathbf{b}}{(\mathbf{a}'\Sigma_{11}\mathbf{a})^{\frac{1}{2}}(\mathbf{b}'\Sigma_{22}\mathbf{b})^{\frac{1}{2}}} \quad (1)$$

CCA では、相関係数 (1) を最大にする  $\mathbf{a}, \mathbf{b}$  を求め

る。すなわち、次の最適化問題 (2) を解く。

$$\begin{aligned} & \underset{\mathbf{a}, \mathbf{b}}{\text{maximize}} && \rho(\mathbf{a}, \mathbf{b}) = \mathbf{a}'\Sigma_{12}\mathbf{b} \\ & \text{subject to} && \mathbf{a}'\Sigma_{11}\mathbf{a} = \mathbf{b}'\Sigma_{22}\mathbf{b} = 1. \end{aligned} \quad (2)$$

$\mathbf{a}, \mathbf{b}$  を正準ベクトル、 $\boldsymbol{\eta}, \boldsymbol{\phi}$  を正準変量、 $\rho$  を正準相関係数と呼ぶ。以下、CCA における各入力データ群をモダリティと呼び、その総数を  $K$  とする。

### 2.2 深層正準相関分析 (DCCA) [1]

DCCA では、多層ニューラルネットワークによって非線形変換されたデータ  $f_1(X), f_2(Y)$  が、(1) を最大にするような  $f_1, f_2$  とを求める。モダリティ  $k$  における第  $l$  層の出力  $\mathbf{h}_l$  は、ユニット数を  $u$ 、活性化関数を  $s$  とすると、 $\mathbf{h}_l = s(W_{l-1}^k \mathbf{h}_{l-1}^k + \mathbf{b}_{l-1}^k) \in \mathbb{R}^u$  である。ここで、 $\mathbf{h}_0^1 = \mathbf{x}, \mathbf{h}_0^2 = \mathbf{y}$  である。出力層のユニット数  $o$  とすると、深さ  $d$  層の DCCA の出力は、

$$f_1(\mathbf{x}) = s(W_d^1 \mathbf{h}_{d-1}^1 + \mathbf{b}_d^1) \in \mathbb{R}^o \quad (3)$$

$$f_2(\mathbf{y}) = s(W_d^2 \mathbf{h}_{d-2}^2 + \mathbf{b}_d^2) \in \mathbb{R}^o \quad (4)$$

と書ける。ここで、 $H_1, H_2$  を以下で定義する。

$$H_1 = f_1(X) = [f_1(\mathbf{x}_1), \dots, f_1(\mathbf{x}_N)] \in \mathbb{R}^{o \times N} \quad (5)$$

$$H_2 = f_2(Y) = [f_2(\mathbf{y}_1), \dots, f_2(\mathbf{y}_N)] \in \mathbb{R}^{o \times N} \quad (6)$$

本節の最初に述べたように、目的関数

$$S = \text{corr}(H_1 W_{CCA}^1, H_2 W_{CCA}^2) \quad (7)$$

を最大化するようなパラメータ  $\boldsymbol{\theta}_k = [W_l^k, \mathbf{b}_l^k]$  を求めることが本手法の目的である。 $H_1, H_2$  をデータ平均について標準化したものを  $\bar{H}_1, \bar{H}_2$  とすると、共分散行列及び交差共分散行列は以下のように計算される。

$$\Sigma_{11} = \frac{1}{N-1} \bar{H}_1 \bar{H}_1^\top + r_1 I \in \mathbb{R}^{o \times o} \quad (8)$$

$$\Sigma_{22} = \frac{1}{N-1} \bar{H}_2 \bar{H}_2^\top + r_2 I \in \mathbb{R}^{o \times o} \quad (9)$$

$$\Sigma_{12} = \frac{1}{N-1} \bar{H}_1 \bar{H}_2^\top \quad (10)$$

Proposal of Deep Generalized Canonical Correlation Analysis and its application for High-accuracy Cross-modal Retrieval

Tomoaki IWASE<sup>†</sup>, Hideki NAKAYAMA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Technology, University of Tokyo

113-8654, 7-3-1, Bunkyo-ku, Tokyo, Japan  
{iwase, nakayama}@nlab.ci.i.u-tokyo.ac.jp

ここで、 $\Sigma_{11}, \Sigma_{22}$  の第 2 項は正則化項であり、 $r_1 > 0, r_2 > 0$  は正則化パラメータである。ここで、 $\Sigma_{11}, \Sigma_{22}$  を正則と仮定すると、相関係数 (7) の大きい方から  $k$  個の和は、行列  $T = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$  の特異値の大きい方から  $k$  個の和に等しい。したがって、 $k = o$  のとき、

$$S = \|T\|_{tr} = \text{tr}\{(T^T T)^{\frac{1}{2}}\} \quad (11)$$

と書ける。また、これを達成する CCA の射影行列は、 $T$  の特異値分解を  $T = UDV^T$  として、

$$W_{CCA}^1 = \Sigma_{11}^{-\frac{1}{2}} U, \quad W_{CCA}^2 = \Sigma_{22}^{-\frac{1}{2}} V \quad (12)$$

である。DCCA のパラメータ  $\theta_k$  は、(11) を用いて、勾配降下法により更新する。このとき、逆誤差伝播に必要な  $H_k$  に対する (11) の変化率、すなわち  $\partial S / \partial H_k$  は以下のように計算される。

$$\frac{\partial S}{\partial H_1} = \frac{1}{N-1} (2\nabla_{11} \bar{H}_1 + \nabla_{12} \bar{H}_2). \quad (13)$$

$$\frac{\partial S}{\partial H_2} = \frac{1}{N-1} (\nabla_{21} \bar{H}_1 + 2\nabla_{22} \bar{H}_2). \quad (14)$$

ただし、

$$\nabla_{11} = -\frac{1}{2} \Sigma_{11}^{-\frac{1}{2}} U D U^T \Sigma_{11}^{-\frac{1}{2}} \quad (15)$$

$$\nabla_{12} = \nabla_{21}^T = \Sigma_{11}^{-\frac{1}{2}} U V^T \Sigma_{22}^{-\frac{1}{2}} \quad (16)$$

$$\nabla_{22} = -\frac{1}{2} \Sigma_{22}^{-\frac{1}{2}} V D V^T \Sigma_{22}^{-\frac{1}{2}} \quad (17)$$

である。以降は通常のニューラルネットワークと同様にして、逆誤差伝播を行う。

### 2.3 一般化正準相関分析 (GCCA)

$X_k \in \mathbb{R}^{N \times p_k}, W_k \in \mathbb{R}^{p_k \times t}, \forall k \in [1, \dots, K]$  をそれぞれ中心化されたモダリティ  $k$  のデータ群とそれに対する射影行列とする。GCCA は、CCA を  $K \geq 2$  に拡張した手法であるが、その定式化は様々存在する [2, 5, 6]。

解析的に解が得られる Carroll ら [2] の規準では、 $X = [X_1, \dots, X_K] \in \mathbb{R}^{N \times p}, W = [W_1^T, \dots, W_K^T]^T \in \mathbb{R}^{p \times t}$  ( $p = \sum_k p_k$ ) とおき、以下の最適化問題を解く。

$$\begin{aligned} \underset{W}{\text{maximize}} \quad & S = \text{tr}(W^T X^T X W) \\ \text{subject to} \quad & W^T D W = I_t \end{aligned} \quad (18)$$

ただし、 $D = \text{blkdiag}(D_1, \dots, D_K)$  はブロック対角行列であり、 $D_K = X_K^T X_K$  である。

### 3 深層一般化正準相関分析 (DGCCA)

DGCCA(図 1) では、DCCA に倣い、多層ニューラルネットワークによって非線形変換された複数のデータ

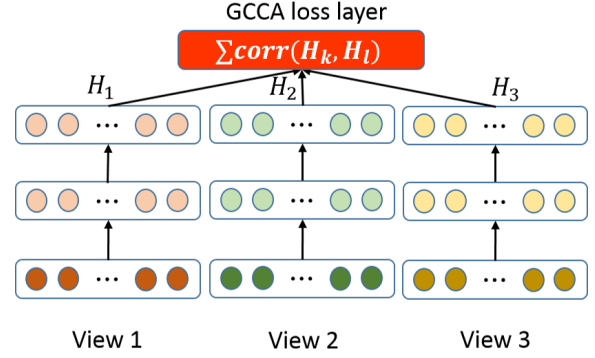


図 1: 提案手法の Deep GCCA モデル ( $M = 3$ )

群に対して、GCCA の規準を最大にするような学習を行う。このとき、DCCA 同様  $\partial S / \partial H_k$  を求める必要がある。例えば (18) においては、次のように計算される。

$$\frac{\partial S}{\partial H} = \left[ \frac{\partial S}{\partial H_1}, \dots, \frac{\partial S}{\partial H_K} \right] = 2XWW^T \quad (19)$$

なお、本手法の応用例としては、クロスモーダル検索の高精度化が挙げられる。Andrew らは複数のデータセットにおいて、CCA や KernelCCA (KCCA) と比較して、DCCA が最も高い交差相関をもつ正準空間を見出した [1]。Gong らは画像とタグにおいてクロスモーダル検索を行い、KCCA と比較して、画像の教師なし分類の結果を第 3 の入力として用いた Kernel GCCA の検索性能が上回ることを示した [3]。これらの知見より、DGCCA は KCCA や DCCA と比べてクロスモーダル検索の性能が上回ることが期待される。

### 参考文献

- [1] G. Andrew et al. Deep canonical correlation analysis. In *Proc. 30th ICML*, pages 1247–1255, 2013.
- [2] J.D. Carroll. A generalization of canonical correlation analysis to three or more sets of variables. In *Proc. 76th APA Annual Convention*, pages 227–228, 1968.
- [3] Y. Gong et al. A multi-view embedding space for modeling internet images, tags, and their semantics. In *Proc. IJCV*, pages 1247–1255, 2013.
- [4] H. Hotelling. Relations between two sets of variates. *Biometrika*, pages 321–377, 1936.
- [5] J.R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

- [6] J. Via et al. Canonical correlation analysis (cca) algorithms for multiple data sets: Application to blind simo equalization. In *Proc. 13th EUSIPCO*, pages 1–4. IEEE, 2005.