1R-07

# Zero-Shot 言語横断文書検索のための画像媒介学習

舟木類佳 † 中山英樹 †

† 東京大学大学院情報理工学研究科

## 1 はじめに

言語横断文書検索（Cross-Lingual Document Retrieval: CLDR）はある言語の文書をクエリとして他言語の関連する文書を見つけ出すタスクである．従来の CLDR における学習は対訳コーパスを必要とするが，一般には十分な対訳コーパスを得ること自体が困難である．

そこで本研究では，対訳コーパスが一切ない（Zero-Shot），あるいは少ししか存在しない場合（Few-Shot）における CLDR のための学習手法を提案する．Web 上の文書には豊富なテキストと画像のペアが存在する．我々の提案する画像媒介学習ではそれらのペアを利用し，十分なテキストペアが存在しない場合でも画像を媒介させることにより 2 つの言語間の関係を間接的に求める．

より具体的には，画像とテキストを含む 2 つの言語のドキュメントに対し，一般化正準相関分析（Generalized Canonical Correlation Analysis: GCCA）の応用により画像をハブとして用いることで共通意味空間を求める．提案手法により対訳コーパスが不足している状態において検索精度を高めることが可能となる．

## 2 関連研究

Rupnik らは十分な対訳の文書が得られる言語（英語等）をハブとして用いている [1] が，介在する言語との対訳コーパスが存在することが前提となっており，それすら得ることができない場合には利用できない．一方で我々は画像をハブとして用いており，学習のためにそれぞれの言語に閉じた文書を利用することができる．殆どの Web 文書は一つの言語に閉じていると考えられ，多くの豊富なマルチメディアデータを内包していることを考えると，我々のセットアップはより利用しやすいものであると思われる．

**Image-Mediated Learning for Zero-Shot Cross-Lingual Document Retrieval**
Ruka FUNAKI†, Hideki NAKAYAMA†
†Graduate School of Information Science and Technology, The University of Tokyo
113-8654, Bunkyo-ku, Tokyo, Japan
{funaki, nakayama}@nlab.ci.i.u-tokyo.ac.jp

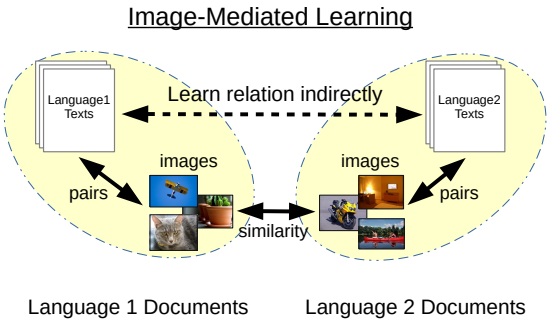


図 1: 画像媒介学習のコンセプト図．互いに似た画像を含む 2 つの異なる言語による文書があった場合，テキストも類似している可能性が高いという考えに基づき間接的に 2 つの言語の関係を学習する．

表 1: 学習データ，及びテストデータの区分

| | 区分 | 英語 | 画像 | 日本語 |
|---|---|---|---|---|
| 1 | [train-E/I] | $E_1$ | $I_1$ | - |
| 2 | [train-I/J] | - | $I_2$ | $J_2$ |
| 3 | [train-E/J] | $E_3$ | - | $J_3$ |
| 4 | [test-E/J] | $E_4$ | - | $J_4$ |

## 3 提案手法

### 3.1 手法の概要

3 種類のデータ，英語，画像，日本語を利用し，それぞれを $E$，$I$，$J$ と表す．表 1 にあるようにデータを重複なく区分する．例えば [train-E/I] は英語と画像から構成される学習用文書を表し，$E_1$ は区分 [train-E/I] における英語テキストを表す．

システムの概要は図 2 のとおりである．データから取られた特徴量を PCA によって圧縮し，GCCA によって学習を行う．テスト時には，特徴量を PCA によって圧縮した上で GCCA により射影し，結合空間上における日本語から英語の最近傍探索を行う．

### 3.2 GCCA による画像媒介学習

3 つのデータの相関の和を最大にするような射影を GCCA[3] で学習する．特徴ベクトル $\mathbf{x}_k, \forall k \in \{E, I, J\}$

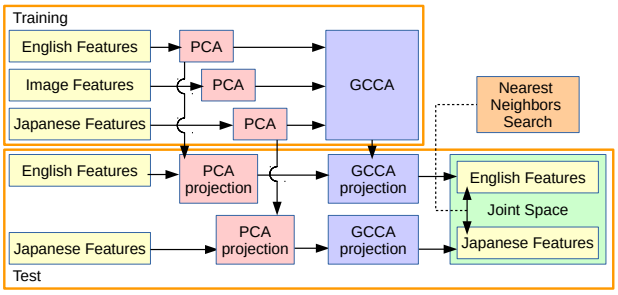


図 2: システムの概要

に対して，正準ベクトル $\mathbf{z}_k = (\mathbf{x}_k - \overline{\mathbf{x}}_k)\mathbf{h}_k$ を求める．射影行列 $\mathbf{h}_k$ は次の一般化固有値問題の解として得られる．

$$\frac{1}{2}\begin{pmatrix} \mathbf{0} & \Sigma_{EI} & \Sigma_{EJ} \\ \Sigma_{IE} & \mathbf{0} & \Sigma_{IJ} \\ \Sigma_{JE} & \Sigma_{JI} & \mathbf{0} \end{pmatrix}\mathbf{h} = \rho\begin{pmatrix} \Sigma_{EE} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{II} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{JJ} \end{pmatrix}\mathbf{h}$$

ここで $\mathbf{h} = (\mathbf{h}_E^T, \mathbf{h}_I^T, \mathbf{h}_J^T)^T$ である．表 1 にある 3 つのモダリティのうち 2 つしか学習データでは存在しない．我々はシンプルに共分散行列を存在するデータのみから計算することによってこの問題を扱う．例えば，Few-Shot 学習においては $\Sigma_{EI}$ は $E_1$ と $I_1$，$\Sigma_{EE}$ は $E_1$ と $E_3$ を用いて計算した．Zero-Shot 学習においては [train-E/J] は存在しないので，$\Sigma_{EE}$ を $E_1$ のみから計算し，$\Sigma_{EJ}$ にはゼロ行列を用いた．

### 3.3 UIUC Pascal Sentence データセット

UIUC Pascal Sentence データセット [2] は 1000 枚の画像と，それに付随した英語で内容を説明した 5 つの文章で構成されている．このデータに我々は日本語の翻訳を加えた [1]．本実験ではそれぞれの画像に対する 5 つの文章をまとめて一つの文書として扱う．

### 3.4 媒介データの数を変化させた精度比較実験

本研究では画像認識における深層学習のフレームワーク Caffe によって提供される ILSVRC2012 データセットで事前学習された畳み込みニューラルネットワークのモデル（GoogLeNet）を利用する．テキストの特徴量として Bag of Words を用いた．日本語の形態素解析に MeCab を用いた．

画像の媒介データのサンプル数を 100〜400 と変化させながら実験を行い，また [train-E/J] の数（横軸）を徐々に 0 から 100 まで変化させながら英語と日本語のみを利用した正準相関分析（CCA）との比較を行った．テストデータを 100 サンプル用い日本語から英語への TOP1 検索の精度に関して評価を行った．Chance Rate は 1%となる．それぞれの実行において，ランダムに置き換えたデータを使った 50 回の試行を行い，結果を平均した．PCA による特徴量の圧縮次元数は 100 次元，最近某探索にはユークリッド距離を用いた．

[1] データセット: http://www.nlab.ci.i.u-tokyo.ac.jp/dataset/pascal_sentence_jp/

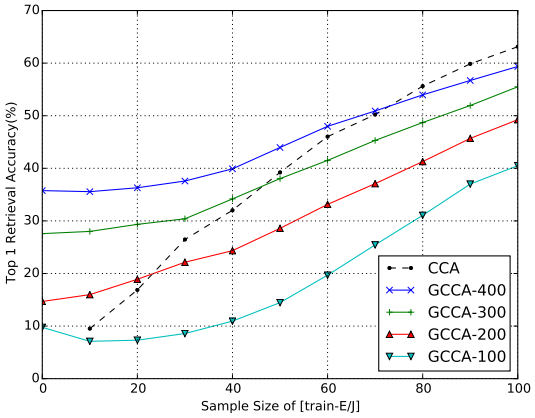


図 3: 実験結果．GCCA-400 は GCCA において [train-E/I] や [train-I/J] がそれぞれ 400 サンプルであることを示す．

実験結果は図 3 のようになった．図から [train-E/J] がゼロの時，最大で 37%のスコアを達成しており Zero-Shot 学習が実現されていることがわかる．また，媒介データが多いほど精度が上がっている．データセットのサイズの制約によりこれ以上媒介データを増やすことができないが，増やすことで媒介学習の精度がより高くなることが期待される．対訳データが少ない場合には画像媒介学習の精度が CCA のベースラインよりも高くなることが分かり，画像媒介学習モデルは Few-Shot 学習においても効果的であることが確認できる． これらの実験により画像媒介学習の有効性が示された．

## 参考文献

[1] Rupnik el al. Cross-Lingual Document Retrieval through Hub Languages. In NIPS Workshop, 2012.

[2] Rashtchian et al. Collecting Image Annotations Using Amazon's Mechanical Turk. In Proc. NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, 2010.

[3] Via et al. Canonical correlation analysis (CCA) algorithms for multiple data sets: Application to blind SIMO equalization. In EUSIPCO, 2005.