2F1-1

潜在分布のカーネル埋め込みによる異種データ間マッチング

Cross-Domain Matching via Kernel Embeddings of Latent Distributions

吉川 友也[*1] 岩田 具治[*2] 澤田 宏[*3] 山田 武士[*2]
Yuya Yoshikawa Tomoharu Iwata Hiroshi Sawada Takeshi Yamada

[*1]奈良先端科学技術大学院大学 Nara Institute of Science and Technology
[*2]NTT コミュニケーション科学基礎研究所 NTT Communication Science Laboratories
[*3]NTT サービスエボリューション研究所 NTT Service Evolution Laboratories

We address a problem of finding matching between different types of data such as Japanese-English documents and images and their text captions in a supervised way. On this problem, we cannot use distance between the different types of data because these data are represented as features in different spaces. We propose a kernel-based matching method that first represents all the data as distributions in a shared latent space, and finds matching between the data based on the distance in the shared latent space. In our experiments, we show that the proposed method outperforms conventional linear and non-linear matching methods on multi-lingual Wikipedia datasets.

1. はじめに

異なる種類のデータ間の適切なマッチングを予測する問題は，自然言語処理や情報検索，データマイニング等で現れる．具体例は，多言語文の対応付け [Zhang 13]，画像と説明文の対応付け [Socher 10]，異なるデータベース間のユーザ情報の対応付け [Li 09] である．本稿では，訓練データとして異種データ間のマッチング情報が与えられる教師あり学習の設定の下で，新しいデータ間のマッチングを予測する問題に取り組む．

入力文書と類似するドキュメントの検索等，同じ種類のデータ間でのマッチング問題では，データ間のマッチングの存在を評価するためにデータ間の距離（もしくは類似度）が利用できる．一方で，異なる種類のデータ間のマッチング問題では，異なる種類のデータは異なる特徴から構成されるため，直接的に距離を測ることはできない．例えば，異なる言語の文書は異なる語彙から構成されるため，異なる言語の文書間の距離を直接測ることはできない．

異種データ間マッチングにおける上記の困難さに対する一つの解決策は，全てのデータを一つの共有の潜在空間へ写像することである．これをするための一つの方法が，正準相関分析 (CCA) [Hotelling 36] である．CCA では，マッチングのあるデータ間の相関が大きくなるように部分空間への線形写像行列を学習する．CCA を適用することにより，異種データ間の距離を測れるようになるが，CCA は線形モデルであるため，実データに現れる複雑なマッチングを予測することは困難である．非線形なマッチングを予測するために，CCA を非線形拡張したカーネル CCA が使える．これまでの研究で，カーネル CCA は多言語文書 [Vinokourov 03] のマッチングや画像とアノテーションのマッチング [Hardoon 06] で良い性能を示したことが報告されている．

カーネル CCA の性能は，同じ種類のデータ間の類似度（カーネル）が正確に定義できるかに依存する．しかし，線形カーネル，多項式カーネル，ガウスカーネル等の多くのカーネル関数は，データを表す特徴ベクトル間の内積に基いているため，2 つのデータに現われる役割が似ている特徴の共起を考慮することができない．例えば，データが文書の場合，「パソコン」と「コンピュータ」は異なる単語であるが類似する物を表す．それにもかかわらず，「パソコン」だけを含む文書と「コンピュータ」だけを含む文書間のカーネルの値は，線形カーネルや多項式カーネルの場合は 0，ガウスカーネルの場合はこれらの特徴ベクトルの長さ (ノルム) だけで決まる．この問題点は，カーネルを利用する SVM やガウス過程回帰においても同様である [Yoshikawa 14, Yoshikawa 15].

本稿では，カーネル CCA における上記の問題点を解決するとともに，識別的にマッチングを予測することができるカーネルに基づく異種データ間マッチングモデルを提案する．図 1に提案法の概要を示す．提案法は，全てのデータに含まれる各特徴を一つの共有空間の潜在ベクトルとして表す．これにより，提案法は既存のカーネル関数の問題点を解決し，異なる特徴間の役割の近さを捉えることができる．その上で，提案法は各

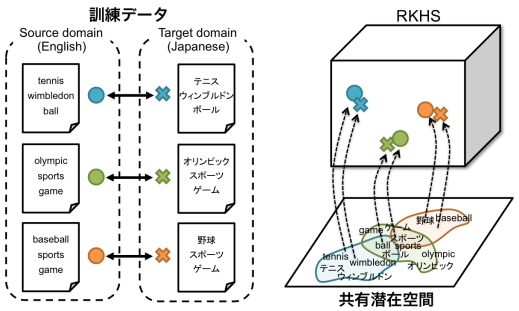

図 1: 多言語文書マッチングの場合の提案法の概要図．二言語のドキュメントペアが訓練データとして観測される．提案法は，各単語（特徴）が共有空間上の潜在ベクトルを持つと仮定する．その上で，各文書はその文書に現れる単語の潜在ベクトルの分布として表現され，分布は分布のカーネル埋め込みに基いて，再生核ヒルベルト空間（RKHS）の点として表現される．ペアの文書間の RKHS 上の点が近づくように特徴の潜在ベクトルを学習した後で，提案法は新しいデータ間のマッチングを発見する．

連絡先: 吉川 友也，奈良先端科学技術大学院大学，yoshikawa.yuya.yl9@is.naist.jp

データをデータに含まれる特徴の潜在ベクトルの分布として表現し，分布間の距離が小さいデータ間にマッチングがあると仮定する．潜在ベクトルの分布と分布間の距離を効率的かつノンパラメトリックに表現するために，分布のモーメント情報を保存できる分布のカーネル埋め込みの枠組みを用いる．提案法の学習では，訓練データ中でマッチングのあるデータ間の潜在分布の距離が小さくなるように，特徴の潜在ベクトルを推定する．推定された潜在ベクトルを使って新しいデータの潜在分布を得ることにより，新しいデータ間のマッチングを予測することが可能である．

実験では，多言語 Wikipedia データセットを使って，提案法が既存モデルに比べて精度良くマッチングを発見できることを示す．

2. 分布のカーネル埋め込み

この節では，提案法で用いられる分布のカーネル埋め込みの枠組みを導入する．分布のカーネル埋め込みは，観測空間 $\mathcal{X}$ 上の任意の確率分布 $\mathbb{P}$ をカーネル k によって定まる再生核ヒルベルト空間 (Reproducing Kernel Hilbert Space; RKHS) $\mathcal{H}_k$ に写像する技術であり，その確率分布は RKHS 上の点 $m^*(\mathbb{P})$ として表現される．具体的には，確率分布 $\mathbb{P}$ が与えられた時，その分布のカーネル埋め込み $m^*(\mathbb{P})$ は，

$$m^*(\mathbb{P}) := \mathbb{E}_{\mathbf{x}\sim\mathbb{P}}[k(\cdot,\mathbf{x})] = \int_{\mathcal{X}} k(\cdot,\mathbf{x})d\mathbb{P} \in \mathcal{H}_k, \tag{1}$$

と定義される．ここで，カーネル k は埋め込みカーネルと呼ばれる．カーネル埋め込み表現 $m^*(\mathbb{P})$ は，カーネル k としてガウスカーネルを使うことにより，確率分布 $\mathbb{P}$ の平均，共分散，さらに高次のモーメント情報を持つことが知られている [Sriperumbudur 09].

現実的には，確率分布 $\mathbb{P}$ は未知で，代わりにその分布からの N 個のサンプルから成る集合 $\mathbf{X} = \{\mathbf{x}_s\}_{s=1}^n$ が観測できる．この場合のカーネル埋め込み表現の推定量は，

$$m(\mathbf{X}) = \frac{1}{n}\sum_{s=1}^{n} k(\cdot,\mathbf{x}_s) \in \mathcal{H}_k, \tag{2}$$

で与えられる．

2.1 分布間の距離計算

式 (2) のカーネル埋め込み表現を使うことにより，二つの分布間の距離を測ることができる．二つのサンプル集合 $\mathbf{X} = \{\mathbf{x}_s\}_{s=1}^n, \mathbf{Y} = \{\mathbf{y}_{s'}\}_{s'=1}^{n'}$ が与えられたとき，式 (2) を適用することにより，これらのカーネル埋め込み表現 $m(\mathbf{X}), m(\mathbf{Y})$ が得られる．その上で，$m(\mathbf{X})$ と $m(\mathbf{Y})$ の距離は，

$$D(\mathbf{X},\mathbf{Y}) \quad = \quad ||m(\mathbf{X}) - m(\mathbf{Y})||^2_{\mathcal{H}_k} \tag{3}$$

で与えられる．直感的には，この距離にはこれらの分布のモーメントがどれだけ異なるかが反映される．また，この距離は，二つの分布の独立性検定等で使われている Maximum Mean Discrepancy (MMD) の二乗と等価である [Gretton 08]．式 (3) は，以下のように展開することにより計算する．

$$\begin{aligned} ||m(\mathbf{X}) - m(\mathbf{Y})||^2_{\mathcal{H}_k} &= \langle m(\mathbf{X}), m(\mathbf{X})\rangle_{\mathcal{H}_k} \\ &+ \quad \langle m(\mathbf{Y}), m(\mathbf{Y})\rangle_{\mathcal{H}_k} - 2\langle m(\mathbf{X}), m(\mathbf{Y})\rangle_{\mathcal{H}_k}, \end{aligned} \tag{4}$$

ここで，$\langle\cdot,\cdot\rangle_{\mathcal{H}_k}$ は RKHS 上の内積を示す．具体的には，$\langle m(\mathbf{X}), m(\mathbf{Y})\rangle_{\mathcal{H}_k}$ は，

$$\begin{aligned} &\langle m(\mathbf{X}), m(\mathbf{Y})\rangle_{\mathcal{H}_k} \\ &= \quad \left\langle \frac{1}{n}\sum_{s=1}^{n} k(\cdot,\mathbf{x}_s), \frac{1}{M}\sum_{s'=1}^{n'} k(\cdot,\mathbf{y}_{s'}) \right\rangle_{\mathcal{H}_k} \\ &= \quad \frac{1}{nn'}\sum_{s=1}^{n}\sum_{s'=1}^{n'} k(\mathbf{x}_s,\mathbf{y}_{s'}). \end{aligned} \tag{5}$$

で与えられる．ここで，$\langle m(\mathbf{X}), m(\mathbf{X})\rangle_{\mathcal{H}_k}, \langle m(\mathbf{Y}), m(\mathbf{Y})\rangle_{\mathcal{H}_k}$ も式 (5) を使い計算できる．

3. 提案法

3.1 モデル

N 個のマッチングのあるデータペアの訓練データ集合 $\mathcal{O} = \{(d_i^s, d_i^t)\}_{i=1}^N$ が与えられるとする．ここで，d_i^s は i 番目の元ドメインのデータ，d_i^t は i 番目の目標ドメインのデータを表す．d_i^s と d_i^t はそれぞれ，元ドメイン特徴集合 $\mathcal{F}^s$ と目標ドメイン特徴集合 $\mathcal{F}^t$ に含まれる特徴の多重集合として表現される．例えば，元ドメインが日本語文書，目標ドメインが英語文書の場合，$\mathcal{F}^s$ は日本語の語彙集合，$\mathcal{F}^t$ は英語の語彙集合である．その場合，d_i^s と d_i^t は，i 番目の文書に含まれる日本語単語と英語単語の集合として表現される．以下では，元ドメインのデータから対応する目標ドメインのデータを見つけることを考える．

提案法は，元ドメイン特徴集合と目標ドメイン特徴集合の各特徴 $f \in \mathcal{F}^s, g \in \mathcal{F}^t$ が，q 次元の共有潜在空間の潜在ベクトル $\mathbf{x}_f, \mathbf{y}_g \in \mathbb{R}^q$ で表現されると仮定する．元ドメイン特徴の潜在ベクトル集合を $\mathbf{X} = \{\mathbf{x}_f\}_{f\in\mathcal{F}^s}$，目標ドメイン特徴の潜在ベクトル集合を $\mathbf{Y} = \{\mathbf{y}_g\}_{g\in\mathcal{F}^t}$ と定義する．その上で，元ドメインと目標ドメインの各データ d_i^s, d_i^t は，データに含まれる特徴の潜在ベクトルの集合 $\mathbf{X}_i = \{\mathbf{x}_f\}_{f\in d_i^s}, \mathbf{Y}_i = \{\mathbf{y}_g\}_{g\in d_i^t}$ として表現される．

2 節では，確率分布をノンパラメトリックに表現し，分布間の距離を計算する方法として，分布のカーネル埋め込みの枠組みを導入した．提案法では，データを表す潜在ベクトルから対応する確率分布を得るために，分布のカーネル埋め込み技術を利用する．まず，元ドメインにおける i 番目のデータと目標ドメインにおける j 番目のデータに対応するカーネル埋め込み表現は，式 (2) から，

$$m(\mathbf{X}_i) = \frac{1}{|\mathbf{X}_i|}\sum_{f\in d_i^s} k(\cdot,\mathbf{x}_f), \;\; m(\mathbf{Y}_j) = \frac{1}{|\mathbf{Y}_j|}\sum_{g\in d_j^t} k(\cdot,\mathbf{y}_g) \tag{6}$$

で与えられる．これらのデータ間の距離は，式 (4) から，

$$D(\mathbf{X}_i,\mathbf{Y}_j) = ||m(\mathbf{X}_i) - m(\mathbf{Y}_j)||^2_{\mathcal{H}_k} \tag{7}$$

で計算できる．

提案法は，マッチングのあるデータ間は似ている潜在ベクトルの分布を持ち，そうでなければ似ていない分布を持つと仮定する．この仮定に基づき，マッチングの尤度を定義する．元ドメインにおける i 番目のデータと目標ドメインにおける j 番目のデータの間にマッチングのある確率 (尤度) を，

$$p(d_j^t | d_i^s, \mathbf{X}, \mathbf{Y}, \theta) = \frac{\exp\left(-D(\mathbf{X}_i,\mathbf{Y}_j)\right)}{\sum_{j'=1}^{N}\exp\left(-D(\mathbf{X}_i,\mathbf{Y}_{j'})\right)} \tag{8}$$

と定義する．ここで，θ は式 (2) で使用する埋め込みカーネルのパラメータである．式 (8) は，元ドメインのデータ d_i^s が与えられた時，目標ドメインのデータ d_j^t が選ばれる確率を表す．

潜在ベクトル $\mathbf{X}, \mathbf{Y}$ の事後確率を定義する．$\mathbf{X}, \mathbf{Y}$ に対して，精度パラメータ $\rho > 0$ の正規分布

$$p(\mathbf{X}|\rho) \propto \prod_{\mathbf{x}\in\mathbf{X}} \exp\left(-\frac{\rho}{2}||\mathbf{x}||_2^2\right), p(\mathbf{Y}|\rho) \propto \prod_{\mathbf{y}\in\mathbf{Y}} \exp\left(-\frac{\rho}{2}||\mathbf{y}||_2^2\right)$$

を事前分布として仮定することにより，事後確率は，

$$p(\mathbf{X}, \mathbf{Y}|\mathcal{O}, \Theta) = \frac{1}{Z} p(\mathbf{X}|\rho) p(\mathbf{Y}|\rho) \prod_{i=1}^{N} p(d_i^t|d_i^s, \mathbf{X}, \mathbf{Y}, \theta) \quad (9)$$

で計算できる．ここで，$\Theta = \{\theta, \rho\}$ はハイパーパラメータ集合，$Z = \int\int p(\mathbf{X}, \mathbf{Y}, \mathcal{O}, \Theta) d\mathbf{X} d\mathbf{Y}$ は周辺尤度で，$\mathbf{X}, \mathbf{Y}$ に対しては定数である．

3.2 学習法

提案法の学習では，事後確率 (9) を最大化することにより，潜在ベクトル $\mathbf{X}, \mathbf{Y}$ の推定を行う．式 (9) の代わりに，以下の負の対数事後確率を考える．

$$\begin{aligned}\mathcal{L}(\mathbf{X}, \mathbf{Y}) &= \sum_{i=1}^{N} D(\mathbf{X}_i, \mathbf{Y}_i) + \log \sum_{j=1}^{N} \exp\left(-D(\mathbf{X}_i, \mathbf{Y}_j)\right) \\ &\quad + \frac{\rho}{2}\left(\sum_{\mathbf{x}\in\mathbf{X}} ||\mathbf{x}||_2^2 + \sum_{\mathbf{y}\in\mathbf{Y}} ||\mathbf{y}||_2^2\right). \end{aligned} \quad (10)$$

そして，式 (10) を最小化するような潜在ベクトル $\mathbf{X}, \mathbf{Y}$ を求める．$\mathbf{X}, \mathbf{Y}$ に関して式 (10) を最小化するために，勾配に基づく最適化を行う．各 $\mathbf{x}_f \in \mathbf{X}$ に関する式 (10) の勾配は，

$$\frac{\partial \mathcal{L}(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{x}_f} = \sum_{i:f\in d_i^s} \frac{\partial D(\mathbf{X}_i, \mathbf{Y}_i)}{\partial \mathbf{x}_f} - \frac{1}{c_i} \sum_{j=1}^{N} e_{ij} \frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{x}_f} + \rho \mathbf{x}_f \quad (11)$$

で与えられる．ここで，

$$e_{ij} = \exp\left(-D(\mathbf{X}_i, \mathbf{Y}_j)\right), \quad c_i = \sum_{j=1}^{N} e_{ij} \quad (12)$$

である．$\frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{x}_f}$ は，$\mathbf{x}_f$ に関する $\mathbf{X}_i$ と $\mathbf{Y}_j$ の距離の勾配で，

$$\begin{aligned}\frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{x}_f} &= \frac{1}{|\mathbf{X}_i|^2} \sum_{l\in d_i^s} \sum_{l'\in d_i^s} \frac{\partial k(\mathbf{x}_l, \mathbf{x}_{l'})}{\partial \mathbf{x}_f} \\ &\quad - \frac{2}{|\mathbf{X}_i||\mathbf{Y}_j|} \sum_{l\in d_i^s} \sum_{g\in d_i^t} \frac{\partial k(\mathbf{x}_l, \mathbf{y}_g)}{\partial \mathbf{x}_f} \end{aligned} \quad (13)$$

で与えられる．もし，$\mathbf{X}_i$ が $\mathbf{x}_f$ を含んでいなければ，この勾配は零ベクトルとなる．$\frac{\partial k(\mathbf{x}_l, \mathbf{x}_{l'})}{\partial \mathbf{x}_f}$ は，$\mathbf{x}_f$ に関する埋め込みカーネル k の勾配である．これは埋め込みカーネルの選択に依存するため，詳細は省略する．

同様に，各 $\mathbf{y}_g \in \mathbf{Y}$ に関する式 (10) の勾配は，

$$\begin{aligned}&\frac{\partial \mathcal{L}(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{y}_g} \\ &= \sum_{i=1}^{N} \frac{\partial D(\mathbf{X}_i, \mathbf{Y}_i)}{\partial \mathbf{y}_g} - \frac{1}{c_i} \sum_{j:g\in d_j^t} e_{ij} \frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{y}_g} + \rho \mathbf{y}_g \end{aligned} \quad (14)$$

表 1: 各手法のハイパーパラメータの範囲

手法	ハイパーパラメータの範囲
Proposed	潜在次元数 $q \in \{8, 10, 12\}$
	カーネルパラメータ $\gamma \in \{10^{-1}, 10^0, \cdots, 10^3\}$
	正則化パラメータ $\rho \in \{0, 10^{-2}, 10^{-1}\}$
KCCA	潜在次元数 $\{10, 20, \cdots, 100\}$
	カーネルパラメータ $\{10^{-3}, 10^{-2}, \cdots, 10^4\}$
	ノイズパラメータ $\{10^{-3}, 10^{-2}, \cdots, 10^0\}$
CCA	潜在次元数 $\{10, 20, \cdots, 100\}$
	ノイズパラメータ $\{10^{-3}, 10^{-2}, \cdots, 10^0\}$
BILDA	潜在トピック数 $\{10, 20, \cdots, 100\}$

で与えられる．ここで，$\frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{y}_g}$ は，$\mathbf{y}_g$ に関する $\mathbf{X}_i$ と $\mathbf{Y}_j$ の距離の勾配で，

$$\begin{aligned}\frac{\partial D(\mathbf{X}_i, \mathbf{Y}_j)}{\partial \mathbf{y}_g} &= \frac{1}{|\mathbf{Y}_j|^2} \sum_{l\in d_j^t} \sum_{l'\in d_j^t} \frac{\partial k(\mathbf{y}_l, \mathbf{y}_{l'})}{\partial \mathbf{y}_g} \\ &\quad - \frac{2}{|\mathbf{X}_i||\mathbf{Y}_j|} \sum_{f\in d_i^s} \sum_{l\in d_i^t} \frac{\partial k(\mathbf{x}_f, \mathbf{y}_l)}{\partial \mathbf{y}_g} \end{aligned} \quad (15)$$

で与えられる．$\frac{\partial k(\mathbf{x}_l, \mathbf{x}_{l'})}{\partial \mathbf{x}_f}$ は，$\mathbf{x}_f$ に関する埋め込みカーネル k の勾配である．これは埋め込みカーネルの選択に依存するため，詳細は省略する．

学習は，勾配 (11), (14) を使い $\mathbf{X}$ と $\mathbf{Y}$ を交互に更新していき，負の対数事後確率 (10) の減少が収束するまで続ける．

3.3 マッチング予測

モデルの学習後，新しいデータ間のマッチングを予測する．元ドメイン特徴集合 $\mathcal{F}_s$ の要素から構成される新しい元ドメインデータ d_{te}^s が与えられた時，そのデータの潜在ベクトル集合を $\mathbf{X}_{\mathrm{te}} = \{\mathbf{x}_f\}_{f\in d_{\mathrm{te}}^s}$ とする．目標ドメイン特徴集合 $\mathcal{F}_t$ の要素から構成される N_{te} 個の目標データ $d_{\mathrm{te},1}^t, d_{\mathrm{te},2}^t, \cdots, d_{\mathrm{te},N_{\mathrm{te}}}^t$ の中で，新しい元ドメインデータがどの目標データにマッチするかを予測するためには，距離関数 (7) を使い，距離 $D(\mathbf{X}_{\mathrm{te}}, \{\mathbf{y}_g\}_{g\in d_{\mathrm{te},j}^t})$ の小さい順に目標データを選べば良い．

4. 実験

六言語 Wikipedia データセットを使い，異なる言語間の Wikipedia 記事のマッチング予測実験を行う．このデータセットは，日本語 (ja)，英語 (en)，ドイツ語 (de)，フランス語 (fr)，イタリア語 (it)，フィンランド語 (fi) の各言語で書かれた 34,024 の Wikipedia 記事から構成されており，言語は異なるが同じ内容が書かれている記事間でマッチングが行われている．このデータセットから ${}_6C_2 = 15$ 通りの二言語記事セットを作る．例えば，de-en と表記される二言語記事セットは，ドイツ語と英語から構成されるデータを表す．実験では，訓練データとして 1,000 記事，開発データとして 100 記事，テストデータとして 100 記事をランダムに抽出し，これを 10 セット生成する．テストの際はテストデータ間のマッチング情報は隠して，記事の内容から正解のマッチングを予測できるか評価する．各データを表す特徴として，その記事に含まれるストップワードや低頻度語以外の単語を使用する．

比較手法として，カーネル CCA (KCCA) [Vinokourov 03], CCA [Hotelling 36], bilingual latent Dirichlet allocation (BILDA) [Iwata 11], 近傍法 (NN) を使用する．近傍法は，新しいデータが与えられた時，同じ言語内の訓練データの最近

表 2: トップ 10 候補に真のマッチングを含む平均確率と標準偏差．太字は，t 検定により有意水準 0.01 で統計的有意に確率が大きいことを示す．

	Proposed	KCCA	CCA	BILDA	NN
de-en	**0.753 (0.059)**	0.573 (0.064)	0.342 (0.056)	0.360 (0.064)	0.153 (0.026)
de-fi	**0.492 (0.031)**	0.406 (0.050)	0.262 (0.035)	0.253 (0.037)	0.177 (0.027)
de-fr	**0.669 (0.042)**	0.542 (0.055)	0.329 (0.032)	0.312 (0.041)	0.174 (0.023)
en-fi	**0.497 (0.025)**	0.404 (0.072)	0.316 (0.046)	0.249 (0.025)	0.160 (0.022)
en-fr	**0.669 (0.030)**	0.540 (0.035)	0.388 (0.052)	0.328 (0.049)	0.157 (0.035)
fi-fr	**0.593 (0.054)**	0.529 (0.061)	0.220 (0.019)	0.251 (0.051)	0.175 (0.038)
it-de	**0.738 (0.038)**	0.615 (0.051)	0.398 (0.040)	0.365 (0.032)	0.168 (0.017)
it-en	**0.762 (0.031)**	0.584 (0.056)	0.351 (0.050)	0.358 (0.038)	0.165 (0.022)
it-fi	**0.546 (0.040)**	0.437 (0.072)	0.317 (0.031)	0.263 (0.050)	0.194 (0.011)
it-fr	**0.713 (0.048)**	0.578 (0.042)	0.388 (0.035)	0.367 (0.053)	0.201 (0.031)
ja-de	**0.749 (0.036)**	0.598 (0.023)	0.350 (0.038)	0.303 (0.042)	0.174 (0.033)
ja-en	**0.805 (0.054)**	0.614 (0.053)	0.347 (0.048)	0.368 (0.035)	0.170 (0.023)
ja-fi	**0.533 (0.037)**	0.413 (0.040)	0.291 (0.031)	0.237 (0.046)	0.181 (0.016)
ja-fr	**0.696 (0.041)**	0.520 (0.043)	0.317 (0.041)	0.309 (0.038)	0.187 (0.023)
ja-it	**0.713 (0.041)**	0.561 (0.048)	0.352 (0.049)	0.335 (0.038)	0.184 (0.029)

傍とマッチングのある別言語の訓練データを選び，そのデータの近傍にあるテストデータをマッチング結果として出力する．表 1 は，提案法と比較手法のハイパーパラメータの一覧を示す．これらのハイパーパラメータの範囲から，開発データを使って最適なハイパーパラメータを探索した．

表 2 は，3.3節のマッチング予測法によって求めた上位 10 個のマッチング候補に正解のマッチングを含む平均確率と標準偏差を示す．この表から，全ての言語の組み合わせにおいて，提案法 (Proposed) は最も精度良く正しいマッチングを予測できることがわかる．

5. おわりに

本稿では，教師あり異種データ間マッチングのためのモデルを提案した．実験では，多言語 Wikipedia データセットを使用し，提案法が最も良い精度で異なる言語間の記事のマッチングを行えることを示した．

謝辞

本研究は JSPS 特別研究員奨励費 (259867) の助成を受けたものです．

参考文献

[Gretton 08] Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A.: A Kernel Statistical Test of Independence, in *Advances in Neural Information Processing Systems*, pp. 1–8 (2008)

[Hardoon 06] Hardoon, D., Saunders, C., Szedmak, S., and Shawe-Taylor, J.: A Correlation Approach for Automatic Image Annotation, in *Advanced Data Mining and Applications*, pp. 681–692 (2006)

[Hotelling 36] Hotelling, H.: Relations Between Two Sets of Variants, *Biometrika*, Vol. 28, pp. 321–377 (1936)

[Iwata 11] Iwata, T., Watanabe, S., and Sawada, H.: Fashion Coordinates Recommender System Using Photographs from Fashion Magazines, in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pp. 2262–2267, AAAI Press (2011)

[Li 09] Li, B., Yang, Q., and Xue, X.: Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model, *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1–8 (2009)

[Smola 07] Smola, A., Gretton, A., Song, L., and Schölkopf, B.: A Hilbert Space Embedding for Distributions, in *Algorithmic Learning Theory* (2007)

[Socher 10] Socher, R. and Fei-Fei, L.: Connecting Modalities: Semi-Supervised Segmentation and Annotation of Images Using Unaligned Text Corpora, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 966–973 (2010)

[Sriperumbudur 09] Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. R. G.: Hilbert Space Embeddings and Metrics on Probability Measures, *The Journal of Machine Learning Research*, Vol. 11, p. 48 (2009)

[Vinokourov 03] Vinokourov, A., Shawe-Taylor, J., and Cristianini, N.: Inferring a Semantic Representation of Text via Cross-Language Correlation Analysis, in *Advances in Neural Information Processing Systems* (2003)

[Yoshikawa 14] Yoshikawa, Y., Iwata, T., and Sawada, H.: Latent Support Measure Machines for Bag-of-Words Data Classification, in *Advances in Neural Information Processing Systems*, pp. 1961–1969 (2014)

[Yoshikawa 15] Yoshikawa, Y., Iwata, T., and Sawada, H.: Nonlinear Regression for Bag-of-Words Data via Gaussian Process Latent Variable Set Model, in *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (2015)

[Zhang 13] Zhang, T., Liu, K., and Zhao, J.: Cross Lingual Entity Linking with Bilingual Topic Model, in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pp. 2218–2224 (2013)