

## ダイバージェンスに基づく NMF を用いた転移学習

木村 圭吾<sup>†1</sup> 吉田 哲也<sup>‡2</sup>

本稿では、トピックグラフに基づく転移学習法を拡張し、一般化 KL(Kullback-Leibler) ダイバージェンスに基づく NMF(Non-negative Matrix Factorization) を用いた転移学習法を提案する。ダイバージェンスを通じた転移学習の確率的な解釈を目指して、フロベニウスノルムに基いてトピックの関係(トピックグラフ)を活用する転移学習法を拡張し、転移学習を一般化 KL ダイバージェンスに基づく最適化問題として定式化する。最適化規準に対する補助関数を定義し、補助関数から最適化アルゴリズムを導出し、その収束性を示す。提案法を文書クラスタリングに適用し、他手法との比較を通じて提案法の有効性を示す。特に、提案法による転移学習を通じてダイバージェンスを用いた場合でも精度向上を実現できることを示す。

### Topic Graph based Transfer Learning via generalized KL divergence based NMF

KEIGO KIMURA <sup>†1</sup> and TETSUYA YOSHIDA <sup>‡2</sup>

We propose a topic graph based transfer learning method based on Non-negative Matrix Factorization (NMF) with generalized Kullback-Leibler (KL) divergence. In this paper we extend the previous NMF based transfer learning method by utilizing generalized KL divergence based NMF so that better probabilistic interpretation can be obtained with the divergence. The proposed method is formalized as the minimization of an objective function under the divergence, and an auxiliary function for the objective function is defined. From the auxiliary function, we derive a learning algorithm with multiplicative update rules, which are guaranteed to converge. The proposed method is evaluated in terms of document clustering over several well-known benchmark datasets. Especially, one drawback of generalized KL divergence based NMF algorithms is performance degradation compared with Frobenius based ones. The experimental results show that, by utilizing the topic graph, the proposed method enables to boost up the performance even with KL divergence based NMF through transfer learning.

### 1. はじめに

あるドメインで学習した知識の効果的な再利用を目指して、転移学習に関する様々な手法が提案されている<sup>12),14),21)</sup>。転移学習とは、ある問題を解くために、別の関連した問題のデータや学習結果を再利用する学習の枠組みであり、知識を転移する際に学習を行う側を元ドメイン、転移された知識を活用する側を目標ドメインと呼ばれる<sup>9),15)</sup>。

本稿では、トピックグラフに基づく転移学習法<sup>14)</sup>を拡張し、一般化 KL(Kullback-Leibler) ダイバージェンスに基づく NMF を用いた転移学習法を提案する。この手法は NMF (Non-negative Matrix Factorization) と呼ばれる手法<sup>2),3),10),11),19)</sup>を用いた学習を行う際、特徴空間の類似性に着目した転移仮説を設定し、NMF を用いて特徴表現を学習し、特徴表現の関係をグラフ構造として表現して正則化に活用するという転移モデルにより転移学習を行う。ダイバージェンスを通じた転移学習の確率的な解釈を目指して、本稿では文献<sup>14)</sup>の手法を一般化 KL ダイバージェンスを用いた場合に拡張し、転移学習をダイバージェンスに基づく最適化問題として定式化する。最適化規準に対する補助関数を定義し、補助関数から最適化アルゴリズムを導出し、その収束性を示す。

提案法を文書クラスタリングにおける転移学習に適用して評価し、他手法との比較を通じて有効性を確認した。特に、NMF において一般化 KL ダイバージェンスを用いた場合はフロベニウスノルムと比較して精度が悪化するという課題があるが、提案法による転移学習を通じてその分を補い精度向上を実現できること確認した。提案法は NMF に基づいているため元ドメイン、目標ドメインにおいてラベルを必要としないという利点があり、また、両ドメインでクラス数も異なっても構わないため適用範囲が広いと考えられる。

2 節では関連研究を紹介し、3 節で提案法の詳細について述べ、4 節で評価実験を報告する。5 節でまとめと今後の展望を述べる。

#### 1.1 準備

本稿では、行列は太字の大文字、ベクトルは太字のイタリック小文字で表記し、 $\mathbf{X}_{ij}$  で行列  $\mathbf{X}$  の第  $ij$  要素を表す。 $\mathbf{X}$  の転置を  $\mathbf{X}^T$  で表す。正則行列  $\mathbf{A}$  に対する逆行列を  $\mathbf{A}^{-1}$  で表す。与えられたデータが表現される空間をデータ空間、写像先の空間を特徴空間と呼ぶ。

<sup>†1</sup> 北海道大学工学部

Faculty of Engineering, Hokkaido University

<sup>‡2</sup> 北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

データ空間を構成するベクトルを属性、特徴空間を構成するベクトルを特徴と呼び、ベクトルの各要素は非負と仮定する。記号  $\simeq$  は近似を表す。

## 2. 関連研究

### 2.1 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF)<sup>11)</sup> とは、与えられた  $p$  個の属性で表現される非負のデータ行列  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}_+^{p \times n}$  ( $n$  はデータ数) を  $\ast 1$ , 特徴数  $q$  のもとで非負の行列  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_q] \in \mathbb{R}_+^{p \times q}$ ,  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}_+^{q \times n}$  の積により

$$\mathbf{X} \simeq \mathbf{UV} \quad \text{ただし, } \mathbf{U}, \mathbf{V} \text{ は非負行列} \quad (1)$$

と近似分解する行列  $\mathbf{U}, \mathbf{V}$  を求める手法である。行列  $\mathbf{U}$  は特徴行列,  $\mathbf{V}$  は係数行列と呼ばれる。NMF により各データ  $\mathbf{x}_i$  は  $\mathbf{u}_1, \dots, \mathbf{u}_q$  の線形結合として近似的に  $\mathbf{U}\mathbf{v}_i$  と表現されることになる。NMF はデータ行列  $\mathbf{X}$  と行列積  $\mathbf{UV}$  との差異の最小化を通じて行列分解を行い, 式 (1) での近似誤差を表現する以下の目的関数の最小化を行う。

$$J_0 = \|\mathbf{X} - \mathbf{UV}\|^2 \quad (2)$$

ここで  $\|\cdot\|$  は行列のノルム (あるいはダイバージェンス) である。

NMF を次元削減法とみなす場合には, 行列  $\mathbf{U}$  の列ベクトルを特徴とみなし, 係数行列  $\mathbf{V}$  を  $\mathbf{U}$  のもとでの低次元データ表現として扱う。データ処理の際には, NMF により求めた低次元表現である  $\mathbf{V}$  に対して既存の手法を適用することが多い。

### 2.2 転移学習

転移学習とは, ある問題を解くために別の関連した問題のデータや学習結果を再利用する学習の枠組みである<sup>9),15)</sup>。知識を転移する際に学習を行う側を元ドメイン, 転移された知識を活用する側を目標ドメインと呼ぶ<sup>9)</sup>。元ドメインでのデータ行列を  $\mathbf{X}_s$ , 目標ドメインでのデータ行列を  $\mathbf{X}_t$  と表記し, 他の行列も同様な添え字で表記する。転移学習には様々な手法が提案されている<sup>12),14),16),21)</sup>, 文献<sup>9)</sup> では両ドメイン間で似ているために転移可能との仮定を転移仮定, その仮定を表現したものを転移モデルと呼んでいる。

### 2.3 トピックグラフに基づく NMF を用いた転移学習

文献<sup>14)</sup> は, NMF では特徴行列  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_q]$  で張られる部分空間においてデータを近似表現することに着目し, NMF を用いた転移学習法 (TNT) を提案した。この手法は特徴空間の類似性に着目した転移仮説を設定し, NMF により行列  $\mathbf{U}$  を学習し, その関係を

グラフ構造として表現し正則化に活用するという転移モデルにより転移学習を行う。この手法ではグラフ構造を正則化項とする目的関数 ( $J_1$ ) の最小化を行い, また, そのグラフラプリアンとしての解釈 (目的関数  $J_2$ ) も示されている。詳細は文献<sup>14)</sup> を参照されたい。

## 3. ダイバージェンスに基づく NMF を用いた転移学習

### 3.1 一般化 KL ダイバージェンスに基づく NMF

2.1 節で述べたように, NMF は行列積  $\mathbf{UV}$  による近似誤差の最小化を通じて行列分解を行う。誤差の測り方に対しては様々な規範が提案されているが, フロベニウスノルム (以下では  $\|\cdot\|_F$  と表記する) と一般化 KL ダイバージェンスが広く用いられている<sup>11)</sup>。後者を用いた場合, 式 (2) に対応する目的関数は下記となる<sup>11)</sup>。

$$J_3 = D(\mathbf{X} \parallel \mathbf{UV}) \quad (3)$$

ここで  $D(\cdot \parallel \cdot)$  は一般化 KL ダイバージェンスであり, 下記で定義される。

$$D(\mathbf{A} \parallel \mathbf{B}) = \sum_{ij} \left( \mathbf{A}_{ij} \log \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ij}} - (\mathbf{A}_{ij} - \mathbf{B}_{ij}) \right) \quad (4)$$

文書クラスタリングなどのテキスト処理においては, フロベニウスノルムに基づく NMF がよく用いられる<sup>19)</sup>, 一般化 KL ダイバージェンスに基づく NMF はあまり用いられていない。この理由として, 4 節で報告するように式 (4) の  $D(\cdot \parallel \cdot)$  を用いた場合は  $\|\cdot\|_F$  を用いた場合と比較して性能が悪化することが挙げられる。

しかし, 精度などの性能悪化という課題はあるが, 一般化 KL ダイバージェンスを用いた場合には確率的な解釈がしやすいという利点がある<sup>1),13)</sup>。本稿ではダイバージェンスを通じた転移学習の確率的な解釈を目指して, フロベニウスノルムに基づく手法を一般化 KL ダイバージェンスに基づく手法に拡張する。

### 3.2 一般化 KL ダイバージェンスへの転移学習法の拡張

以下では表記を簡略化し, 目標ドメインを表す添え字  $t$  を省略して表記する。元ドメインで学習した特徴行列, および元ドメインでのトピックグラフとそのグラフラプリアンは目標ドメインでの学習の際には定数であることに注意されたい。

式 (4) の  $D(\cdot \parallel \cdot)$  を用いて, 以下の目的関数の最小化を考える。

$$J_4 = \sum_{i,j} \left( \mathbf{X}_{ij} \log \frac{\mathbf{X}_{ij}}{\mathbf{Y}_{ij}} - (\mathbf{X}_{ij} - \mathbf{Y}_{ij}) \right) + \frac{\nu}{2} \sum_{k,s} \sum_i \left( \mathbf{U}_{ik} \log \frac{\mathbf{U}_{ik}}{\mathbf{U}_{is}} + \mathbf{U}_{is} \log \frac{\mathbf{U}_{is}}{\mathbf{U}_{ik}} \right) \mathbf{W}_{ks} \quad (5)$$

<sup>\ast 1</sup> 本稿では, NMF における記法に従い,  $\mathbf{X} \in \mathbb{R}_+^{p \times n}$  の形式をデータ行列と呼ぶ。

ここで  $\mathbf{Y}=\mathbf{UV}$  であり,  $\nu$  は正則化パラメータである. 第二項における  $\mathbf{W}$  は元ドメインで構築されたトピックグラフを表し, これを目標ドメインに転移した式 (5) の最小化を行う. このため, 提案法は元ドメインでのトピックグラフに基づく正則化項を用いた転移学習を行うことになる.

以下では, 可読性のために行列の要素  $\mathbf{X}_{ij}$  を小文字を用いて  $x_{ij}$  とも表記する. Jensen の不等式<sup>4)</sup> に基づき, 式 (5) の目的関数に対する補助関数<sup>3),11)</sup> を定義する \*1.

$$G(\mathbf{U}, \mathbf{U}^{(\tau)}) = \sum_{i,j} \left( x_{ij} \log x_{ij} - x_{ij} + \sum_k u_{ik} v_{kj} \right) - \sum_{i,j,k} \left( x_{ij} \frac{u_{ik}^{(\tau)} v_{kj}}{\sum_k u_{ik}^{(\tau)} v_{kj}} (\log u_{ik} v_{kj} - \log \frac{u_{ik}^{(\tau)} v_{kj}}{\sum_k u_{ik}^{(\tau)} v_{kj}}) \right) + \frac{\nu}{2} \sum_{k,s} \sum_i \left( u_{ik} \log \frac{u_{ik}}{u_{is}} + u_{is} \log \frac{u_{is}}{u_{ik}} \right) w_{ks} \quad (6)$$

$\mathbf{U}^{(\tau)}$  は  $\tau$  回目の更新時での行列  $\mathbf{U}$  を表す. 式 (6) の  $\mathbf{U}$  に関する偏微分により下記を得る.

$$\frac{\partial G(\mathbf{U}, \mathbf{U}^{(\tau)})}{\partial \mathbf{U}} = \sum_j v_{kj} - \sum_j x_{ij} \frac{u_{ik}^{(\tau)} v_{kj}}{\sum_k u_{ik}^{(\tau)} v_{kj}} \frac{1}{u_{ik}} + \frac{\nu}{2} \sum_s (\log \frac{u_{ik}}{u_{is}} + 1 - \frac{u_{is}}{u_{ik}}) w_{ks} \quad (7)$$

$x \rightarrow 1$  で  $\log x \simeq 1 - 1/x$  という近似に基づき<sup>3)</sup>, 式 (7) を以下で近似する.

$$\frac{\partial G(\mathbf{U}, \mathbf{U}^{(\tau)})}{\partial \mathbf{U}} \simeq \sum_j v_{kj} - \sum_j x_{ij} \frac{u_{ik}^{(\tau)} v_{kj}}{\sum_k u_{ik}^{(\tau)} v_{kj}} \frac{1}{u_{ik}} + \frac{\nu}{u_{ik}} \sum_s (u_{ik} - u_{is}) w_{ks} \quad (8)$$

式 (8) の停留条件から以下の更新式を得る.

$$\tilde{\mathbf{u}}_i^{(\tau+1)} = \left[ u_{i1}^{(\tau)} \sum_j \frac{x_{ij} v_{1j}}{\sum_k u_{ik}^{(\tau)} v_{kj}}, \dots, u_{iq}^{(\tau)} \sum_j \frac{x_{ij} v_{qj}}{\sum_k u_{ik}^{(\tau)} v_{kj}} \right] (\text{diag}(\sum_j v_{kj}) + \nu \mathbf{L})^{-1} \quad (9)$$

ここで  $\tilde{\mathbf{u}}_i$  は行列  $\mathbf{U}$  の  $i$  番目の行ベクトルであり,  $\mathbf{L}$  は元ドメインにおけるトピックグラフに対するグラフラプラシアンである. このため, 元ドメインで構築したトピックグラフのグラフラプラシアンを用いた更新を行うことになる. また,  $\text{diag}(\sum_j v_{kj}) \in \mathbb{R}_+^{q \times q}$  は各  $k$  に

対する和をそれぞれ対角要素とする対角行列であり, 以下で定義される.

$$\text{diag}(\sum_j v_{kj}) = \text{diag}(\sum_j v_{1j}, \dots, \sum_j v_{qj}) \quad (10)$$

式 (5) で行列  $\mathbf{V}$  に依存するのは第一項のみのため,  $\mathbf{V}$  の更新式は文献<sup>11)</sup> と同じとなる.

上記の更新式を用いる提案アルゴリズムについて以下の性質が成り立つ.

**定理 1.** 式 (5) はアルゴリズム KITNT の更新式により単調非増加であり, 非負であるためアルゴリズムは収束する.

収束性の証明は紙面の都合から省略する.

### 3.3 目標ドメインのデータ対に基づく正則化

2.2 節で述べたように転移学習とは目標ドメインでの性能向上を目指すものである. 目標ドメインのデータをより反映した係数行列  $\mathbf{V}_t$  を学習するために, データ対に基づく正則化項<sup>3)</sup> を追加した以下の目的関数の最小化を考える.

$$J_5 = J_4 + \frac{\lambda}{2} \sum_{j,s} \sum_k \left( v_{kj} \log \frac{v_{kj}}{v_{ks}} + v_{ks} \log \frac{v_{ks}}{v_{kj}} \right) w'_{js} \quad (11)$$

式 (11) の第二項は文献<sup>3)</sup> で提案されたデータ対に基づく正則化項に対応し,  $\lambda$  はパラメータである. 文献<sup>3)</sup> ではデータ空間において  $m$ -近傍グラフを構築し \*2, その隣接行列を用いるが, 提案法では文献<sup>14)</sup> と同様に重み行列 (式 (11) では  $w'_{js}$  と表記) を用いる.

式 (11) の最小化に対しても, 3.2 節と同様な更新式の導出が可能である. 特に, 式 (11) の行列  $\mathbf{U}$  と  $\mathbf{V}$  に関する偏微分は独立なため行列  $\mathbf{U}$  の更新式は不変であり, 式 (9) と同じ更新式となる. 他方, 行列  $\mathbf{V}$  の更新式は文献<sup>3)</sup> と同じとなる. さらに, 独立性により式 (11) を用いる場合でも定理 1 と同様にアルゴリズムの収束性を示すことができる.

## 4. 評価

### 4.1 実験設定

#### 4.1.1 対象データ

先行研究<sup>12),14),20)</sup> に基づき, 提案法を単語頻度に基づくベクトル空間モデルで表現された文書データである, 1) 20 ニュースグループ (以下, 20NG と表記) \*3, 2) SRAA \*4, に

\*2 各頂点ごとに類似度の降順に  $m$  個の他の頂点と接続して構築される.

\*3 <http://people.csail.mit.edu/~jrennie/20Newsgroups/> (本稿では 20news-18828 を使用)

\*4 <http://www.cs.umass.edu/~mccallum/data.html>

\*1 補助関数の定義については付録で述べる.

表 1 20 Newsgroup における階層  
Table 1 hierarchy in 20 Newsgroup dataset

トップ カテゴリ	サブ カテゴリ	表記	トップ カテゴリ	サブ カテゴリ	表記
comp	graphics	comp-1	sci	crypt	sci-1
	os.ms.windows.misc	comp-2		electronics	sci-2
	sys.ibm.pc.hardware	comp-3		med	sci-3
	sys.mac.hardware	comp-4		space	sci-4
	windows.x	comp-5			
rec	autos	rec-1	talk	politics.guns	talk-1
	motorcycles	rec-2		politics.mideast	talk-2
	sport.baseball	rec-3		politics.misc	talk-3
	sport.hockey	rec-4		religion.misc	talk-4

表 2 4 clusters dataset

Table 2 4 clusters dataset

id	Dataset	Clusters					
		Source	comp-1	comp-4	rec-2	rec-4	
a	comp vs. rec	comp-2	comp-5	rec-1	rec-3		
		Target	rec-3	rec-1	rec-3		
b	comp vs. sci	Source	comp-1	comp-2	sci-1	sci-2	
		Target	comp-4	comp-5	sci-3	space	
c	comp vs. talk	Source	comp-1	comp-5	talk-2	talk-4	
		Target	comp-2	comp-3	talk-1	talk-3	
d	rec vs. talk	Source	rec-1	rec-2	talk-1	talk-3	
		Target	rec-3	rec-4	talk-2	talk-4	
e	sci vs. talk	Source	sci-2	sci-3	talk-3	talk-4	
		Target	sci-1	sci-4	talk-1	talk-2	

表 3 6 clusters dataset

Table 3 6 clusters dataset

id	Dataset	Clusters							
		Source	comp-1	comp-2	rec-1	rec-2	sci-1	sci-2	
f	comp vs. rec vs. sci	comp-3	comp-4	rec-3	rec-4	sci-3	sci-4		
		Target	comp-3	comp-4	rec-3	rec-4	sci-3	sci-4	
g	comp vs. rec vs. talk	Source	comp-1	comp-2	rec-1	rec-2	talk-1	talk-2	
		Target	comp-3	comp-4	rec-3	rec-4	talk-3	talk-4	
h	comp vs. sci vs. talk	Source	comp-1	comp-2	sci-1	sci-2	talk-1	talk-2	
		Target	comp-3	comp-4	sci-3	sci-4	talk-3	talk-4	
i	rec vs. sci vs. talk	Source	rec-1	rec-2	sci-1	sci-2	talk-1	talk-2	
		Target	rec-3	rec-4	sci-3	sci-4	talk-3	talk-4	

適用して文書クラスタリングに対する転移学習を行い、従来法との比較を行った。

20NG にはトップカテゴリとして 4 つのトップカテゴリ {comp,rec,sci,talk} があり、その下にサブカテゴリがある (表 1 参照)。両ドメインでトップカテゴリは共通とし、ドメイン同士は異なるサブカテゴリに属すデータセットを作成した。転移学習はデータ量の多い元ドメインから知識を学習してデータ量の少ない目標ドメインでの性能向上を目指すものであるため、元ドメインと目標ドメインとのデータ数の比を 4:1 とし、目標ドメインでの各サブカテゴリから 25 個の文書を非復元抽出し、各データセットごとに 10 個ずつサンプルを作成した。データセットを表 2, 表 3 に示す。なお、転移学習では両ドメインでのラベ

表 4 SRAA データセット

Table 4 SRAA dataset

id	Dataset	Clusters		
		Source	sim-auto	real-auto
j	auto vs. aviation	Target	sim-aviation	real-aviation
		Source	real-auto	real-aviation
k	real vs. sim	Target	sim-auto	sim-aviation
		Source	sim-auto	sim-aviation

ルあり・なしの設定があるが<sup>9),15)</sup>、提案法は NMF に基づいているため両ドメインともにラベルは不要である。このため、以下では両ドメインともにラベルを用いない、教師なし転移学習に対する実験を行った。

SRAA は UseNet における 4 つのカテゴリでの投稿記事を集めたものである。73,218 の記事から本文を抜き出し、短文の記事を除外するために各カテゴリごとにファイルサイズで上位 500 個の記事を抽出し、20NG と同様な処理を行い 10 個のサンプルを作成した。実験に使用したデータセットを表 4 に示す。

各サンプルごとに porter stemmer <sup>\*1</sup> を用いて stemming を行い、MontyTagger <sup>\*2</sup> を用いて品詞に分解し、stop word を除去して相互情報量で上位 2,000 語の単語を選択した。

#### 4.1.2 評価尺度

上記のデータは、各データ (ここでは文書) ごとに真のクラスタが既知である。各データセットに対して、各データに対する真のクラスタと割り当てられたクラスタに基づいて正規化相互情報量 ( $NMI$ ) を評価した。真のクラスタと割り当てられたクラスタに対応する確率変数を  $C, \hat{C}$  とすると、正規化相互情報量 ( $NMI$ ) は以下で定義される。

$$NMI = \frac{I(\hat{C}; C)}{(H(\hat{C}) + H(C))/2} \quad (\in [0, 1]) \quad (12)$$

$H(\cdot)$  はシャノン情報量であり、 $I(\cdot; \cdot)$  は相互情報量である。 $NMI$  における正規化には様々な手法があるが<sup>17)</sup> 本稿では算術平均を用いた。 $NMI$  が大きいほど真のクラスタでのデータ割当てに合致することを示すため、クラスタ割当ての精度に対応する。実験では目標ドメインにおける  $NMI$  を評価した。

#### 4.1.3 比較手法

代表的な NMF の手法として、1) NMF<sup>11)</sup>, 2) WNMF<sup>19)</sup>, 3) LPNMF<sup>3)</sup> に対し提案法による転移学習の効果を評価した。また、他の転移学習法として、2.2 節で述べた 4) SDT<sup>12)</sup>, 5)

\*1 <http://www.tartarus.org/~martin/PorterStemmer>

\*2 <http://web.media.mit.edu/~hugo/montytagger>

MTrick<sup>21)</sup>と比較した。更に、文書クラスタリングに対するベースラインとして6) skmeans<sup>5)</sup>を評価した。

WNMFとは、グラフラプシアンを用いる Ncut 法における重み付け<sup>18)</sup>を活用してデータ行列を変換し、変換後のデータ行列に対して NMF を適用する手法である<sup>19)</sup>。

LPNMF<sup>3)</sup>とは、3.3節で述べたように式(3)と式(11)の第二項を用いる手法であり、フロベニウスノルム ( $\|\cdot\|_F$ ) に基づく GNMF<sup>2)</sup>を式(4)の  $D(\cdot\|\cdot)$  に拡張したものである。

4.1.4 実験手順

各データセットにおいてトップカテゴリに対するクラスタリングを行った。トップカテゴリの数をクラスタ数  $k^*1$  とし、特徴数  $q$  をクラスタ数  $k$  に設定して構築した係数行列  $\mathbf{V}$  に対し、クラスタ数  $k$  のもとで skmeans 法を適用しクラスタリングを行った。NMF は局所最適化を行う手法であるため、得られる結果(行列  $\mathbf{U}, \mathbf{V}$ ) は初期値に依存する。このため、ひとつのデータ行列に対して行列の初期値(図 ??では2行目)をランダムに10回変えた。さらに、各データセットで10個のサンプルに同様な処理を行い計100回試行の平均を求めた。その際、NMFでの交互最適化終了条件は繰返し数30とした。

また、表2, 3, 4においてSource(元ドメイン), Target(目標ドメイン)と表記しているが、この役割を交換した場合<sup>\*2)</sup>の実験も行い、結果(NMI)を平均した。

4.1.5 実験パラメータ

データ間の類似度としては、文書処理で標準的に用いられるコサイン類似度を用いた。SDT, MTrickでは元ドメインでラベルが与えられることを仮定しているが、評価実験ではラベルは用いないため、ラベルに基づく正則化項に対する係数を0に設定した。

文献<sup>12)</sup>に従い、SDTでは目標ドメインに対する重み係数を0.025とし、Ncut<sup>18)</sup>と同様に kmeans を適用した。また、文献<sup>21)</sup>に従い MTrick では元ドメインにおけるワードクラスタ数を50、目標ドメインにおける重み係数を1.5とした。なお、文献<sup>21)</sup>のようにラベルに基づいてロジスティック回帰を用いた初期値設定ができないため、2.2節の行列  $\mathbf{G}$  の初期値は kmeans を用い、行列  $\mathbf{F}$  は文献<sup>21)</sup>に従い pLSI<sup>8)</sup> を用いて初期値を設定した。

LPNMF<sup>3)</sup> (および GNMF<sup>2)</sup>) に対しては各データごとに類似度が上位10個の近傍データを用いて  $m$ -近傍グラフを構築し、式(11)での  $\lambda$  はそれぞれ文献<sup>2),3)</sup>で推奨された値( $\lambda=100$ )とした。他の手法と同様に本稿の手法もパラメータ値に依存するが、文献<sup>14)</sup>にあ

\*1 表2では2クラスタ, 表3では3クラスタ表4では2クラスタとなる。  
\*2 ドメインの役割のみを交換し、データ数の比などは4.1.1節での記述に従いデータセットを作成した。

表5 Comparison (NMI, with skmeans)  
Table 5 Performance of NMF algorithms (NMI)

	method	a	b	c	d	e	f	g	h	i	j	k
$\ \cdot\ _F$	NMF	0.416	0.263	0.648	0.373	0.144	0.334	0.377	0.322	0.244	0.760	0.396
	WNMF	0.162	0.135	0.423	0.228	0.111	0.266	0.288	0.250	0.194	0.331	0.083
	GNMF	0.261	0.114	0.568	0.252	0.078	0.265	0.307	0.250	0.194	0.525	0.393
$D(\cdot\ \cdot)$	NMF	0.071	0.076	0.129	0.088	0.045	0.097	0.106	0.104	0.077	0.198	0.139
	WNMF	0.064	0.061	0.116	0.115	0.045	0.095	0.106	0.104	0.077	0.192	0.127
	LPNMF	0.184	0.181	0.464	0.418	0.080	0.230	0.265	0.237	0.123	0.414	0.240
	skmeans	0.208	0.107	0.213	0.143	0.154	0.206	0.224	0.208	0.185	0.331	0.394

わせて  $\nu=0.15, \lambda=1.5$  とした。

4.2 結果

4.2.1 フロベニウスノルムとダイバージェンスの比較

フロベニウスノルム ( $\|\cdot\|_F$ ) に基づく NMF と一般化 KL ダイバージェンス(式(4)の  $D(\cdot\|\cdot)$ ) に基づく NMF の比較を表5に示す。3.1節で述べたように、表5で対応する手法を比較すると<sup>\*3)</sup> 近似誤差の測り方の違いにより後者の方が前者より精度(NMI)が非常に悪化している。このため、ダイバージェンスに基づく NMF は転移学習の確率的な解釈につながるも期待されるが、精度の観点からは課題があることがわかる。

4.2.2 他手法との比較

他の転移学習法との比較を表6に示す<sup>\*4)</sup>。以下では、表中で本稿での提案法(kITNT)は+kIT, 文献<sup>14)</sup>の手法(TNT)は+Tと示した。

表6の結果より、提案法は他の転移学習法(SDT, MTrick)を多くのデータセットで上回った。特に、式(11)に基づく提案法(LPnMF+kIT)はMTrickを多少上回る性能を示した。NMF+kITとWNMF+kITはMTrickには及ばなかったが、1つのデータセットを除いてSDTを上回った。

また、トピックグラフに基づく転移学習法における近似誤差の測り方( $\|\cdot\|_F, D(\cdot\|\cdot)$ )の影響についても、一般化KLダイバージェンスを用いることにより本稿での提案法がフロベニウスノルムを用いる TNT を上回る性能を示した。

5. おわりに

本稿では、トピックグラフに基づく転移学習法<sup>14)</sup>を拡張し、一般化KL(Kullback-Leibler)

\*3 たとえば、 $\|\cdot\|_F$ を用いる NMF と  $D(\cdot\|\cdot)$ を用いる NMF の比較。  
\*4 データセットごとに最良の結果(NMI)を太字とした。

表 6 Comparison (NMI, with skmeans)  
Table 6 Comparison of transfer learning methods (NMI)

	method	a	b	c	d	e	f	g	h	i	j	k
TNT ( $\ \cdot\ _F$ )	NMF+T	0.540	0.335	0.747	0.366	0.141	0.334	0.404	0.332	0.210	<b>0.801</b>	0.384
	WNMF+T	0.352	0.232	0.544	0.290	0.138	0.292	0.371	0.303	0.183	0.454	0.108
	GNMF+T	<b>0.581</b>	0.355	0.755	0.442	0.166	<b>0.376</b>	0.477	0.386	0.250	0.696	<b>0.393</b>
klTNT ( $D(\cdot\ \cdot)$ )	NMF+klT	0.372	0.399	0.807	0.373	0.192	0.295	0.477	0.386	0.281	0.470	0.374
	WNMF+klT	0.359	0.402	0.797	0.355	0.200	0.285	0.479	0.383	0.271	0.452	0.351
other transfer	LPNMF+klT	0.417	<b>0.444</b>	<b>0.826</b>	<b>0.461</b>	<b>0.236</b>	0.346	<b>0.540</b>	<b>0.419</b>	<b>0.326</b>	0.414	0.240
	MTrick	0.425	0.255	0.718	0.437	0.164	0.374	0.433	0.348	0.248	0.764	0.289
	SDT	0.233	0.140	0.409	0.159	0.047	0.096	0.172	0.198	0.083	0.763	0.189

ダイバージェンスに基づく NMF(Non-negative Matrix Factorization) を用いた転移学習法を提案した。ダイバージェンスを通じた転移学習の確率的な解釈を目指して、本稿ではトピックの関係 (トピックグラフ) に基づく転移メカニズムを拡張し、一般化 KL ダイバージェンスに基づく最適化問題として定式化した。最適化規準に対する補助関数を定義し、補助関数から最適化アルゴリズムを導出し、その収束性を示した。

提案法を文書クラスタリングにおける転移学習に適用して評価し、提案法による転移学習の効果を確認した。特に、NMF において一般化 KL ダイバージェンスを用いた場合はフロベニウスノルムと比較して精度が悪化するという課題があるが、提案法による転移学習を通じてその分を補い精度向上を実現できること確認した。今後はダイバージェンスに基づいて情報幾何などの転移学習の確率的な解釈<sup>1),13)</sup> に取り組んでいきたい。

### 参 考 文 献

- 1) ichi Amari, S.: Information Geometry of the EM and em Algorithms for Neural Networks, *Neural Networks*, Vol.8, pp.1379–1408 (1995).
- 2) Cai, D., He, X., Wu, X. and Han, J.: Non-negative Matrix Factorization on Manifold, *Proc. ICDM'08*, pp.63–72 (2008).
- 3) Cai, D., He, X., Wang, X., Bao, H. and Han, J.: Locality preserving nonnegative matrix factorization, *Proc. IJCAI'09*, pp.1010–1015 (2009).
- 4) Cover, T. and Thomas, J.: *Elements of Information Theory*, Wiley (2006).
- 5) Dhillon, J. and Modha, D.: Concept decompositions for large sparse text data using clustering, *Machine Learning*, Vol.42, pp.143–175 (2001).
- 6) Ding, C., Li, T., Peng, W. and Park, H.: Orthogonal Nonnegative Matrix Tri-factorizations for Clustering, *Proc. KDD '06*, pp.126–135 (2006).
- 7) Harville, D.A.: *Matrix Algebra From a Statistician's Perspective*, Springer (2008).
- 8) Hofmann, T.: Probabilistic Latent Semantic Indexing, *Proc. of SIGIR'99*, pp.50–57 (1999).
- 9) 神鷹敏弘: 転移学習, 人工知能学会誌, Vol.25, No.4, pp.572–580 (2010).

- 10) Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788–791 (1999).
- 11) Lee, D.D. and Seung, H.S.: Algorithms for Non-negative Matrix Factorization, *Proc. NIPS'01*, pp.556–562 (2001).
- 12) Ling, X., Dai, W., Xue, G., Yang, Q. and Yu, Y.: Spectral Domain-Transfer Learning, *Proc. KDD'08*, pp.488–496 (2008).
- 13) Murata, N., Takenouchi, T., Kanamori, T. and Eguchi, S.: Information Geometry of U-Boost and Bregman Divergence, *Neural Networks*, Vol.16, pp.1432–1481 (2004).
- 14) 荻野広樹, 吉田哲也: トピックグラフに基づく NMF を用いた転移学習, 情報処理学会論文誌:数理モデル化と応用, Vol.4, No.3, pp.73–83 (2011).
- 15) Pan, S.J. and Yang, Q.: A Survey on Transfer Learning, *IEEE Transactions on Knowledge and Data Engineering*, Vol.22, No.10, pp.1345–1359 (2010).
- 16) S., B.-D., Blitzer, J., K., C. and Pereira, F.: Analysis of representations for domain adaptation, *Proc. NIPS'07*, pp.137–144 (2007).
- 17) Strehl, A. and Ghosh, J.: Cluster Ensembles -A Knowledge Reuse Framework for Combining Multiple Partitions, *J. Machine Learning Research*, Vol.3, No.3, pp. 583–617 (2002).
- 18) von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol.17, No.4, pp.395–416 (2007).
- 19) Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *Proc. SIGIR'03*, pp.267–273 (2003).
- 20) Xue, G., Dai, W., Yang, Q. and Yu, Y.: Topic-bridged PLSA for Cross-Domain Text Classification, *Proc. SIGIR'08*, pp.627–634 (2008).
- 21) Zhuang, F., Luo, P., Xiaogang, H., He, Q., Xiong, Y. and Shi, Z.: Exploiting Associations between Word Clusters and Document Classes for Cross-domain Text Categorization, *Proc. SDM'10*, pp.13–24 (2010).