対訳辞書構築技術に基づく教師なしデータフュージョン

熊谷雄介 [1]　野沢悠哉 [2]　横井祥 [3,4]　道本龍 [1]
[1] 株式会社博報堂 DY ホールディングス　[2] 株式会社博報堂 DY メディアパートナーズ
[3] 東北大学　[4] 理研 AIP

{yusuke.kumagae, ryo.domoto}@hakuhodo.co.jp
yuya.nozawa@hakuhodody-media.co.jp　yokoi@tohoku.ac.jp

概要

データフュージョンは複数のテーブルデータを擬似的につなぎ合わせる情報推薦分野のタスクで，現実的な教師なしの設定ではまだ解くことができていない．本論文では，このタスクと自然言語処理分野の対訳辞書構築タスクが「空間を超えてベクトルの類似性が一貫する」という共通点を持つことに着目し，対訳辞書構築技術による教師なしデータフュージョンをおこなう．複数の実データでの実験結果とあわせて報告する．

1　はじめに

データフュージョン [1, 2] は複数のテーブルデータを擬似的につなぎ合わせる技術で，情報推薦分野におけるデータ拡張のアプローチのひとつである．たとえば「商品購買履歴テーブル」と「Web サイト閲覧履歴テーブル」の各行（ユーザ）をおおよそでも対応づけられれば，店頭でホットサンドメーカーを買ったであろうユーザにソロキャンプ動画を推薦するといった，単一データからでは実現できないリッチな施策が可能になる．データフュージョンはこのようなマーケティング分野 [3] や，症状や治療法に関する疎なテーブル同士を統合して知見を得たい医療分野 [4, 5] において盛んに活用されている．

データフュージョンのためのこれまでの手法は，二つのテーブルの間で一部の行（ユーザ）もしくは列（特徴量）が事前に対応付けられていることを仮定しており，利用可能な情報に応じて様々な手法が提案されている（表 1）．

しかし現実の問題においては，ユーザ（行）や特徴量（列）について事前には対応付けがわからない場合が多い．ユーザ（行）についてはプライバシー保護や技術規制 [6] による利用難度の高まり，特徴量（列）については新たな知見を得るという目的からしてそもそも二つのテーブルには共通する項目が存在しない，という理由が挙げられる[1)]．

本論文では，二つのテーブルデータにおいてユーザと変数のいずれもが共通して存在しない，より現実的な設定でのデータフュージョンを**教師なしデータフュージョン**と呼び，これに取り組む（表 1）．

本論文では，データフュージョンの問題と教師なし対訳辞書構築 (bilingual lexicon induction) の問題が形式的に非常に近いことに着目し，その手法を応用する．対訳辞書構築は，複数言語間の単語の対応関係を推定する問題で，最近は教師なしの手法が盛んに開発されている [10, 11, 12, 13, 14]．一見するとこの二つの問題は全く異なる分野の異なる問題に見えるが，表 2 に示すような対応関係が存在する．詳細な比較は 3.1 章で行う．

本論文の貢献は (1) より現実的な設定のデータフュージョンである教師なしデータフュージョンを定式化し，はじめてこれに取り組んだ，(2) 教師なしデータフュージョンと教師なし対訳辞書構築が同じ形式をとることに着目し，Gromov-Wasserstein 距離に基づく対訳辞書構築技術を用いた手法を提案した，(3) 複数の実データを用いた実験により，ベースラインに対する優位性を示したの三つである．

2　問題設定

教師なしデータフュージョンは二つのテーブルデータ（**ソース**と**ターゲット**）について，対応する行を教師なしで対応付けるタスクである[2)]．

1)　冒頭の例でも一方の列は商品，もう一方の列は Web サイトであり，クラスからして異なる．
2)　テーブルとは例えば購買データやテレビの視聴行動データであり，各行がそれぞれのユーザの行動履歴（購買データではどの商品を購入したのか，視聴行動データでは誰がどの時間帯にテレビを視聴していたかなど）である．

表 1　ユーザ（行）および特徴量（列）が利用可能な時に採用できる手法の一覧.

	共通列あり	共通列なし
共通行あり	-	Orthogonal Procrustes [7], CDMCA [8]
共通行なし	マハラノビスマッチング [9]，傾向スコア [4]	本論文

表 2　対訳辞書構築とデータフュージョンの対応関係.

	対訳辞書構築	教師なしデータフュージョン
二つの空間	異言語 (英語とスペイン語など)	異種データ (購買行動と視聴行動など)
各空間の表現	単語埋め込み行列	ユーザ・アイテム行列
対応づけたい対象	単語ベクトル (行)	ユーザベクトル (行)
空間の特徴	行の類似性は空間を超えて一貫	行の類似性は空間を超えて一貫
教師データ	部分的な対応 or **教師なし**	部分的な対応 or **教師なし**

ソース，ターゲットをそれぞれ $\mathbf{T}_s \in \mathbb{R}^{n\times m}, \mathbf{T}_t \in \mathbb{R}^{n'\times m'}$ と表す．n, n' はソース，ターゲットに含まれるユーザ数であり，m, m' はソース，ターゲットの各ユーザの特徴量ベクトルの長さである．

正解の対応関係 $\mathbf{Y} \in \{0,1\}^{n\times n'}$ は $y_{i,j} = 1$ の時にはソースの i 行目とターゲットの j 行目が同一人物であり，$y_{i,j} = 0$ の時には別人であることを意味する．教師なしデーアフュージョンの目的はソースとターゲットの対応の推定であり，ソースの i 行目とターゲットの j 行目が対応する度合い $\mathbf{\Pi}_{i,j}$ を要素に持つ行列 $\mathbf{\Pi} \in \mathbb{R}^{n\times n'}$ の推定である．

3　提案手法

本論文では，教師なしデータフュージョンの問題が教師なし対訳辞書構築の問題と形式的に似ていることに着目し，対訳辞書構築で用いられた Gromov–Wasserstein 距離 (GW) を活用して教師なしデータフュージョンを行う．

3.1　対訳辞書構築とデータフュージョン

本論文で取り組む教師なしデータフュージョンと対訳辞書構築との対応関係を表 2 に示す．まず入出力に関して，異なる空間に存在するベクトル同士を結びつける[3] という同一の形式を取っている．

またいずれの問題も**ベクトル同士の相対的な位置関係は空間を超えて一貫する**と仮定できる．以下これについて詳しく述べる．まず対訳辞書構築のための多くの手法が「異言語にて，各言語の埋め込み空間内での単語の相対的な位置関係は類似している」という仮定を利用している [10, 12][4]．

3)　例えば対訳辞書構築は英語の cat ベクトルと日本語の猫ベクトルとを，データフュージョンはある人の購買ベクトルと別のある人の Web サイト閲覧ベクトルを結びつけるのが目的である．

4)　Mikolov らは猫，牛，犬，馬，豚という五つの動物の単語埋め込みの相対的な位置関係が，英語（ゲルマン語派）とスペイン語（ロマンス諸語）において非常に類似することを示している [15].

我々は教師なしデータフュージョンにおいてもこの仮定が自然に成り立つと考える．例えば購買データとテレビ番組視聴データにおいては，新しいものが好きなユーザはどちらのデータでも新しいアイテム（新商品や新番組）を好むだろう．その結果，どちらのデータにおいても同じ嗜好（新しいもの好き）のユーザと類似した行動を取り，反対に保守的なユーザとは類似しなくなるだろう．このようにして，ユーザの潜在的な興味の近さが複数のテーブルでのユーザベクトル（行）の近さとして観測されると自然に考えられる．

以上の考察から，「相対的位置関係一貫」仮説に基づく教師なし対訳辞書構築技術を教師なしデータフュージョンに活用すること提案する．とくに，(1) 教師なしの状況で適用可能 (2) 学習が安定する の二点を満たす唯一の手法である Gromov–Wasserstein 距離ベースの手法 [11] を利用する．

ただしここで，単語ベクトルは**密**な埋め込みである一方，情報推薦で用いるユーザ・アイテム行列における行（ユーザベクトル）は**疎**であることに注意する．我々は，単語ベクトルが疎な共起行列を分解した結果と解釈できることに注目し [16]，同様に疎なユーザ・アイテム行列を分解して対訳辞書構築技法を適用する[5]．

3.2　Gromov-Wasserstein 距離

本論文では Gromov-Wasserstein 距離 (**GW**) [17] にもとづく対訳辞書構築 [11] を応用する．GW は，二つのデータ $\mathbf{D}$, $\mathbf{D}'$ に含まれる点を対応付ける技術であり，$\mathbf{D}$ において近い二点と $\mathbf{D}'$ において近い二点とを強く，$\mathbf{D}$ において遠い二点と $\mathbf{D}'$ において遠い二点とを弱く対応付ける．GW の計算の詳細は

5)　実験では生の疎なユーザベクトルを用いる場合と行列分解に基づく密ベクトルを用いる場合を比較し，実際に単語ベクトルと設定が近くなる密なベクトルを用いる方が良好な性能を示すことがわかった (5.1 章).

Alvarez-Melis らの論文 [11] を参照されたし．

3.3 提案法

提案法は次のステップで構成される．

・ソース $\mathbf{T}_s$ およびターゲット $\mathbf{T}_t$ からそれぞれのユーザベクトル $\mathbf{V}_s, \mathbf{V}_t$ を求める

・ユーザベクトル $\mathbf{V}_s, \mathbf{V}_t$ からそれぞれの要素間距離行列 $\mathbf{C}_s, \mathbf{C}_t$ を得る

・$\mathbf{C}_s$ および $\mathbf{C}_t$ を用いて GW を求め，輸送計画行列 $\mathbf{\Pi}$ を得，この要素 $\mathbf{\Pi}_{i,j}$ をソースのユーザ i がターゲットのユーザ j に対応する度合いとする

$\mathbf{C}_s, \mathbf{C}_t$ はユーザベクトル $\mathbf{V}_s, \mathbf{V}_t$ それぞれの行間の距離を要素に持つ行列であり，対訳辞書構築では行ベクトルの比較にコサイン距離を用いるのが一般的である．実際に採用した手法は 5.1 章で説明する．

4 関連研究

対訳辞書構築には大きく分けて真に対応関係にある単語ペアを用いる教師ありの手法と，それらを用いない教師なしの手法があるが，本論文では後者に言及する．教師なしの手法は大きく分けて敵対的生成ネットワーク (Generative Adversarial Network, GAN) [18] を用いる手法 [10, 19, 12, 20] と GAN を用いない手法 [11, 21, 22, 23] がある．

GAN の学習がハイパーパラメタに非常に敏感で不安定である [24] 一方，行列積の繰り返しによって収束が保証されている [17] ことから，本論文では GW による手法 [11] を採用した．

5 実験

提案法が実データでうまく働くかを確認するために，**(1) 単一のデータを擬似的に分割したデータ対**および **(2) 異なる種類のデータ対**それぞれに対する教師なしデータフュージョンを行う．

5.1 実験設定

データ　実験には以下のデータを用いた．**Movie** はユーザが映画を 5 段階で評価したデータであり，各行がユーザ，各列が映画のテーブルデータである．これは情報推薦タスクの検証において最も用いられている MovieLens 1M [25] から作成した．**POS** はユーザごとの商品購入情報であり，各行がユーザ，各列が JAN コード単位で区別された商品からなるテーブルデータである[6]．**TV** はユーザごとのテレビ視聴情報であり，各行がユーザ，各列が「月，曜日，時間帯，放送局」[7]における視聴状況からなるテーブルデータである．

POS と TV は同一のユーザが両データに存在するシングルソースデータである[8]．

本実験では章冒頭で述べた通り (1) 単一のデータを擬似的に分割したデータ対と (2) 異なる種類のデータ対との二種類を準備した．以下，前者の分割について説明する[9]．

■分割による擬似的なデータ対の作成：もともとひとつだったテーブルを分割して擬似的に教師なしの問題を作成する．具体的には，各テーブルの列（特徴量）をランダムで 2 クラスにわけ，このクラスに応じて元のテーブルをソースとターゲットに分割する．以上の処理を Movie，POS，TV に実施した上で，二つのテーブルの間の行（ユーザ）の対応関係を忘れて教師なしのアラインメントに取り組む．

■開発データとテストデータの作成：また，ハイパーパラメタ（次元数，距離関数，前処理など）決定のために，テーブルデータを行を基準にランダムに二分割し，一方は開発データ (dev) として最も良いハイパーパラメタを探索するために用い，もう一方はテストデータ (test) として本論文で精度を報告する．データの詳細を表 3 に記す[10]．

評価指標　教師なしデータフュージョンは $\{0, 1\}$ のラベルを実数値で予測する二値分類問題であるので，二値分類問題の標準的な評価尺度である ROC-AUC (Area Under the Receiver Operating Characteristic Curve) [26] を採用する．ソース側の各行に対する予測に対して ROC-AUC を計算し，その平均 (macro ROC-AUC) を報告する[11]．ROC-AUC は $[0, 1]$ の間の値をとり，P(正例のスコア > 負例のスコア) と解釈できる．

ハイパーパラメタ調整　開発データを用いて決定した主なハイパーパラメタは以下である[12]．

ユーザベクトルの構築は，Randomized SVD [27] (SVD) と Alternating Least Squares による Matrix Factorization [28] (ALS) とによって 10 次元に削減したベクトルの二種類をユーザベクトルとた．要素内距

6） POS には主要な商品の JAN コードのみが含まれている．

7） 「12 月，水曜日，21 時台，放送局 X」など

8） 株式会社インテージのインテージ i-SSP のうち，2019 年 1 月 1 日から 2019 年 12 月 31 日までのデータ．

9） 後者は POS と TV である．

10） 開発データの件数も含めた詳細は Appendix に記載した．

11） 異なる五種の乱数で初期化したユーザベクトルによる macro ROC-AUC の平均値と標準偏差を報告する．

12） その他は Appendix を参考されたし．

表 3 実験に用いたデータの統計情報．1 行目は（ソース → ターゲット）の組み合わせを，2 行目と 3 行目はテーブルサイズ（ユーザ数 × 特徴数）を表す．

データ対	ソース	ターゲット
Movie→Movie	3020 ×1811	3020 × 1805
TV → TV	1865 × 17099	1865 × 17095
POS → POS	1865 × 32480	1865 × 32365
TV → POS	1865 × 64845	1865 × 34194
POS → TV	1865 × 34194	1865 × 64845

表 4 擬似的なデータ対における macro ROC-AUC とその標準偏差．値は大きいほど予測精度が高い．太字は各データの最大値．

手法	埋込	Movie	POS	TV
random		.50	.50	.50
cos	raw	-	-	-
	SVD	.73 ±.04	.73 ±.05	**1.00 ±.00**
	ALS	.61 ±.06	.54 ±.07	.55 ±.06
GW	raw	.92	0.72	1.00
	SVD	**.99 ±.00**	**.91 ±.01**	**1.00 ±.00**
	ALS	.93 ±.07	.90 ±.02	.72 ±.27

表 5 異なるデータ対における macro ROC-AUC とその標準偏差．値は大きいほど予測精度が高い．太字は各データの最大値．

手法	埋込	POS → TV	POS → TV
random		.50	.50
cos	raw	-	-
	SVD	.49 ±.01	.49 ±.01
	ALS	**.52 ±.02**	**.52 ±.01**
GW	raw	0.51	0.50
	SVD	.50 ±.00	.50 ±.00
	ALS	**.52 ±.02**	.51 ±.01

離行列の構築ではユーザベクトル $\mathbf{v}_i, \mathbf{v}_j$ に対するコサイン距離 $1 - \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{||\mathbf{v}_i||_2 ||\mathbf{v}_j||_2}$ を用いた．

ベースライン　教師なしデータフュージョンには既存研究や簡単に比較できる手法が存在しないため，いくつかの簡単なベースラインを用意した．

random：ソースに対してランダムにターゲットを対応付けたものを予測結果とする．

cos：**「特徴量空間の違いを無視し，ユーザベクトルのコサイン類似度が近いほど類似したユーザである」**という仮定から，ユーザ i, j の特徴量ベクトル $\mathbf{v}_i$, $\mathbf{v}_j$ のコサイン類似度 $\frac{\mathbf{v}_i \cdot \mathbf{v}_j}{||\mathbf{v}_i||_2 ||\mathbf{v}_j||_2}$ を対応度合いとする．特徴量空間の違いが無視するため，精度良く対応付けられないと考えられる．

5.2 実験結果

擬似的なデータ対に対するデータフュージョン　単一データを分割した擬似データでの結果は表 4 の通り．一切教師情報を与えない難しい問題にも関わらず，提案法（GW）はチャンスレベル（random）を大きく凌駕した．これらの実験で用いたのは疑似テーブル対であり，「行の類似性は空間を超えて一貫」仮説をおおよそ満たす状態であったため，同じ仮説に基づく対訳辞書構築手法の適用というアプローチがうまく動いたと考えらえれる．また，疎な行情報（raw）を使うよりも特に SVD を介した密ベクトルを用いた方が性能が高い傾向が見られた．単語ベクトルの学習は共起行列の SVD と等価であること [16] から，ここでも対訳辞書構築との共通点が分かる．ふたつのタスクの共通点や相違点のより厳密な分析は今後の課題である．

TV では，cos / SVD という一見弱いベースラインが ROC-AUC で 1.0 を示している（全問正解）．この原因は，TV 行列をランダムで擬似分割したとしてもソースとターゲットのテーブルにほとんど「共通」の列（特徴量）が多数残ってたためと考えられる[13]．今後教師なしデータフュージョンのベンチマークを作成する際は，特徴量間の依存性を減らす工夫が必要だと考えられる．

異なるデータ対に対するデータフュージョン　異なるデータに対するデータフュージョンの結果は表 5 の通り．一部チャンスレベルを超える結果は観測できるが，まだとても「解けた」とは言い難い状況である．今後の展望として，まず仮説「行の類似性は空間を超えて一貫」が現実の異種データでどの程度満たされているかについて，データ自体を定性的・定量的に観察したい．また，異種データの場合は対応付かない行[14]の存在が想定される．現状の最適輸送に基づく提案法も，ROC-AUC の平均という評価尺度も，すべての行に対応先があることを仮定しており，手法や評価尺度の改定も今後の重要な仕事であろう．

6 結論

本論文では，共通する行および列を持たない二つのテーブルデータを対応付けるタスクである教師なしデータフュージョンを定式化した．そして教師なしデータフュージョンが教師なし対訳辞書構築が同じ形式をとることに着目し，Gromov-Wasserstein 距離にもとづく対訳辞書構築技術を用いた手法を提案した．複数の実データを用いた検証実験の結果，提案法がベースラインに対して優位性を持つことと，その課題を報告した．今後はより多くの種類のテーブルデータに対する検証を進めたい．

13） 具体的には，「9:00 にテレビ局 A を視聴」と「10:00 にテレビ局 A を視聴」のような列対など．

14） ターゲット側に似たユーザが存在しないようなユーザ

参考文献

[1] Wagner A. Kamakura and Michel Wedel. Statistical data fusion for cross-tabulation. **Journal of Marketing Research**, Vol. 34, No. 4, pp. 485–498, 1997.

[2] Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward, 2010.

[3] Zvi Gilula, Robert E. Mcculloch, and Peter E. Rossi. A direct approach to data fusion. **Journal of Marketing Research**, Vol. 43, No. 1, pp. 73–83, oct 2006.

[4] Paul R. Rosenbaum and Donald B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. **American Statistician**, Vol. 39, No. 1, pp. 33–38, 1985.

[5] Noa Dagan, Noam Barda, Eldad Kepten, Oren Miron, Shay Perchik, Mark A. Katz, Miguel A. Hernán, Marc Lipsitch, Ben Reis, and Ran D. Balicer. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. **New England Journal of Medicine**, Vol. 384, No. 15, pp. 1412–1423, 2021.

[6] Webkit Org. Intelligent Tracking Prevention 2.3, 2021.

[7] Samuel L. Smith, David H.P. Turban, Steven Hamblin, and Nils Y. Hammerla. Offline bilingual word vectors, orthogonal transformations and the inverted softmax, 2017.

[8] Hidetoshi Shimodaira. Cross-validation of matching correlation analysis by resampling matching weights. **Neural Networks**, Vol. 75, pp. 126–140, 2016.

[9] Marcello. D'Orazio, Marco. Di Zio, and Mauro. Scanu. **Statistical Matching: Theory and Practice**. Wiley, 2006.

[10] Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data. In **6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings**, feb 2018.

[11] David Alvarez-Melis and Tommi S. Jaakkola. Gromov-wasserstein alignment of word embedding spaces. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 1881–1890, 2020.

[12] Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. Earth mover's distance minimization for unsupervised bilingual lexicon induction. In **EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings**, pp. 1934–1945, 2017.

[13] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In **ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)**, Vol. 1, pp. 789–798, 2018.

[14] Yedid Hoshen and Lior Wolf. Non-adversarial unsupervised word translation. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 469–478, 2020.

[15] Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. Exploiting Similarities among Languages for Machine Translation. sep 2013.

[16] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In **Advances in Neural Information Processing Systems**, Vol. 3, pp. 2177–2185, 2014.

[17] Gabriel Peyre, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In **33rd International Conference on Machine Learning, ICML 2016**, Vol. 6, pp. 3927–3935, 2016.

[18] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In **Advances in Neural Information Processing Systems**, Vol. 3, pp. 2672–2680, 2014.

[19] Xilun Chen and Claire Cardie. Unsupervised multilingual word embeddings. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 261–270, 2020.

[20] Zi Yi Dou, Zhi Hao Zhou, and Shujian Huang. Unsupervised bilingual lexicon induction via latent variable models. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 621–626, 2020.

[21] Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with Wasserstein Procrustes. In **AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics**, 2020.

[22] Ruochen Xu, Yiming Yang, Naoki Otani, and Yuexin Wu. Unsupervised cross-lingual transfer of word embedding spaces. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 2465–2474, 2020.

[23] Tanmoy Mukherjee, Makoto Yamada, and Timothy Hospedales. Learning unsupervised word translations without adversaries. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018**, pp. 627–632, 2020.

[24] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In **Advances in Neural Information Processing Systems**, pp. 2234–2242, 2016.

[25] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. **ACM Transactions on Interactive Intelligent Systems**, Vol. 5, No. 4, dec 2015.

[26] Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. **Pattern Recognition**, Vol. 30, No. 7, pp. 1145–1159, 1997.

[27] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. **SIAM Review**, Vol. 53, No. 2, pp. 217–288, 2011.

[28] Gábor Takács, István Pilászy, and Domonkos Tikk. Applications of the conjugate gradient method for implicit feedback collaborative filtering. In **RecSys'11 - Proceedings of the 5th ACM Conference on Recommender Systems**, pp. 297–300, 2011.

表 6 実験に用いたデータの統計情報．1 行目は（ソース → ターゲット）の組み合わせを，3 行目と 4 行目はテーブルサイズ（ユーザ数 × 特徴数）を表す．

データ対		ソース	ターゲット
Movie→Movie	dev	3020 × 1817	3020 × 1805
	test	3020 ×1811	3020 × 1805
TV → TV	dev	1865 × 17113	1865 × 17119
	test	1865 × 17099	1865 × 17095
POS → POS	dev	1865 × 32469	1865 × 32525
	test	1865 × 32480	1865 × 32365
TV → POS	dev	1865 × 64994	1865 × 34232
	test	1865 × 64845	1865 × 34194
POS → TV	dev	1865 × 34232	1865 × 64994
	test	1865 × 34194	1865 × 64845

A 実験

ハイパーパラメタ調整

開発データを用いて以下のハイパーパラメタを決定した．

データの表現 データは値が含まれる部分を 1，値が含まれない部分を 0 として二値化した．

ユーザベクトルの構築 Randomized SVD [27][15] (SVD) と Alternating Least Squares による Matrix Factorization [28][16](ALS) とによって 10 次元に削減したベクトルの二種類をユーザベクトルとして用意した．ALS の正則化項は 0.01 を用いた．

ユーザベクトルの後処理 ユーザベクトルは $\mu_{s,j}$ を $\mathbf{V}_s[:,j]$ の平均，$\sigma_{s,j}$ を $\mathbf{V}_s[:,j]$ の標準偏差として標準化 $\mathbf{V}'_s[:,j] = \frac{\mathbf{V}_s[:,j]-\mu_{s,j}}{\sigma_{s,j}}$ を行った ($\mathbf{V}_t$ も同様に実施した)．

要素内距離 要素内距離行列の構築ではユーザベクトル間の距離に二人のユーザ (i,j) のユーザベクトル $\mathbf{v}_i, \mathbf{v}_j$ に対するコサイン距離 $1-\frac{\mathbf{v}_i \cdot \mathbf{v}_j}{||\mathbf{v}_i||_2||\mathbf{v}_j||_2}$ を用いた．

GW の正則化項 Gromov-Wasserstein 距離の計算における正則化項 λ は 0.001 を用いた．

サンプルサイズ

開発データも含めたサンプルサイズの図を表 6 に記す．

15） https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html を用いた
16） https://github.com/benfred/implicit を用いた