

# 半教師あり非負値行列因子分解を用いた二項分類向け転移学習の評価

米川慧<sup>†1</sup> 黒川茂莉<sup>†1</sup> 村松茂樹<sup>†1</sup> 小林亜令<sup>†1</sup>

**概要:** 近年、大規模なデータの蓄積・処理を可能とする手法やツールが成熟してきており、企業や組織によるデータ活用の間口が広がり、新たな価値創造が期待されている。データ活用に際しては、事業領域によっては取得が困難な属性があり、有用な分析が困難となることがある。この問題に対しては、より多くの属性を有する転移元ドメインにおいて構築された非負値行列因子分解の因子を、より少ない属性を有する転移目標ドメインに持ち込み、二つのドメインに共通した属性を手がかりとすることで未知の属性値の推定が可能である。また、転移元ドメインにおける因子の構築に際してラベルデータを用いることで、因子を介してラベルデータの情報を転移目標ドメインに転移することが可能であり、転移目標ドメインにおけるタスクの精度向上が可能だと考えられる。本稿では、転移目標ドメインにおけるタスクの精度改善に向け、Web アクセスログを特徴量とした二項分類タスクを対象に、半教師あり非負値行列因子分解を用いた二項分類向け転移学習の精度を評価した。

**キーワード:** 半教師あり非負値行列因子分解, 転移学習, ロバスト学習, 非負値最小二乗法

## 1. はじめに

ビッグデータと呼ばれる大規模データの活用プロセスは一般に収集・蓄積・処理に分けられるが、近年、どの工程に関してもコモディティ化が進み、企業はデータ活用による価値創造により大きな比重をかけられるようになった。データの収集については、kafka[1]といったメッセージ処理フレームワークや Fluentd[2]といったログコレクタソフトウェアが一般的になっており、大規模データの収集が容易となっている。データの蓄積については、HDD や SSD といった情報記録媒体の容量単価が低減しただけでなく、これらのハードウェアを多数束ねて仮想的に一つの大きな記憶領域としてユーザに提示する HDFS[3]を始めとする分散ファイルシステムにより大規模データの蓄積が容易となっている。データの処理については、Hadoop[4]、Spark[5]といった分散処理フレームワークにより大規模データの取扱いが容易となっただけでなく、MLlib[6]などの機械学習ライブラリや、Theano[7]、TensorFlow[8]といった機械学習の実装を指向した数値計算フレームワークの登場により、既存手法の活用に加え、新規に考案した手法の実装も容易となった。

データ活用を事業に取り込んだ例としては、EC における商品の推薦[9]、VOD や映画などのコンテンツの推薦[10][11]、ニュースメディアにおける記事の推薦 [12]、位置情報を用いた交通量推定[13]などが挙げられる。このような状況を受けて、消費者の利便性向上、顧客獲得効率化、業務プロセス効率化、新規ビジネス創出などのためにデータ活用を実践しようとする企業は少なくないと思われるが、データ活用に際して最初の障壁となり、かつ、最終的に創造される価値の潜在的な総量を決定するのがデータ収集である。

データ活用に際してその目的に見合うデータを収集することは、レコード数と属性の観点から必ずしも容易ではない。データ収集におけるレコード数の問題は、統計的機械学習理論においては VC 次元[14]として学習に必要な十分なサンプル数の下限が示されており、また、レコメンドシステム分野においてはコールドスタート問題[15]として推薦精度が低下する要因として知られている。サブスクライプ型サービスの登録情報や Web アクセスログを始めとする行動履歴を収集し活用するのに十分な量蓄積することは、ある程度の期間待つことが必要となるため、データ活用の取り組みを遅らせる要因となる。しかしながら、時間が解決するという側面がある。

一方、データ収集における属性の問題は、企業の事業領域に応じて取得が容易な属性と取得が困難な属性が存在するという問題である。例えば、EC プラットフォームの運営者は、ユーザによる各出店者のページへのアクセス履歴や商品の購買履歴を収集可能だと考えられ、商品カテゴリ毎に購買経験の有無を示すような属性を網羅することができる。一方、出店者は、自身のページに関するデータを EC プラットフォーム運営者と共有することは考えられるものの、他の出店者に関するデータを共有することは考えにくい。PC 周辺機器を扱う出店者がユーザのプロファイリングを行おうとした場合は PC の購入履歴を参照することが望ましいと考えられるが、PC を扱う出店者のデータを取得できないため有効なプロファイリングができない、という問題が生じる。レコード数の問題と比較すると、属性の問題は時間により解決されるものではないため、解決には企業間での情報連携の手段を講じる必要がある。

企業間情報連携による属性の拡大を実現する手段としては、以下の方法が考えられる。

<sup>†1</sup> (株)KDDI 総合研究所  
KDDI Research, Inc.

## (1) ユーザ ID 連携された生データの共同利用による属性拡大方法

本方法は、企業間においてユーザ ID を連携可能とした上で、各企業において収集・蓄積した生データを転送しユーザ ID により結合することで、一方の企業では取得不可能であった未知属性を他方の企業データから得るといものである。生データを転送するため、精密な分析が可能となる反面、パーソナルデータの取り扱いの課題がある。パーソナルデータを適切に取り扱うため、利用規約の整備、ユーザからの利用許諾の取得、データ管理に関するシステム整備等の準備が必要になり、準備コストがかかる。

## (2) 統計モデル化されたデータの共通属性を用いた属性値の推定による属性拡大方法

本方法は、各企業のデータの共通属性と未知属性の関係を統計モデル化し、統計モデルを互いに連携し、統計モデルと自社データを用いて未知属性値を推定するというものである。未知属性が推定値となるため、ユーザ ID 連携された生データの共同利用による方法と比較して精密な分析はできないが、生データの共同利用に伴うコストを抑えることができる。

事業環境の変化が速く激しい昨今においては、準備コストが相対的に小さく初動に優位性のある(2)の推定による属性拡大方法が望ましいと考えられる。本論文では、推定による属性拡大方法として、より多くの属性を有する転移元ドメインにおいて構築された非負値行列因子分解の因子を、より少ない属性を有する転移目標ドメインに持ち込み、二つのドメインに共通した属性を手がかりとすることで未知の属性値を推定する手法を評価する。本論文の貢献は以下2点である。

- 転移目標ドメインにおける二項分類を目的とし、転移元ドメインにおける非負値行列因子分解に教師ラベルを含めることの効果を検証する。
- 転移目標ドメイン、転移元ドメインのデータを、属性の分布に関してロングテールな特徴を持つ Web アクセスログと想定し、ロングテールの裾部分の情報を損なうことなく因子に反映するのに適した損失関数について実験的に評価する。

以降、2章では関連研究を述べ、3章では問題設定とアプローチを述べ、4章では実験と考察を述べ、5章ではまとめを述べる。

## 2. 関連研究

企業間情報連携による属性拡大という問題設定は、転移学習の一種である。転移学習とは、転移元ドメインからの知識を転移目標ドメインでのタスクに適用し精度を向上させる学習手法[16]であり、[16]の転移学習の分類に従うと本論文は特徴ベース・分離型に相当する。

### ● 特徴ベース転移学習：

特徴ベース転移学習では、転移元ドメインのデータの空間から転移目標ドメインでも有用な部分空間を求める。本論文では、転移元ドメインが転移目標ドメインより多くの属性を有する前提で、転移目標ドメインにおいても転移元ドメイン固有の属性を利用できるように、転移目標ドメインにおいて属性を推定するための有用な部分空間を求める。

### ● 分離型転移学習：

分離型転移学習では、転移元ドメインで転移目標ドメインに適合するようにデータ変換を行い、転移目標ドメインでは転送された変換後データを用いて一般の学習手法により学習を行う。本論文では、生データは転送できないという制約の元、転移元ドメインで属性推定に有用な部分空間を統計モデル化し、統計モデルを転移目標ドメインに転送する。

以降、2.1節では、転移学習に関する応用研究について述べ、本論文での問題意識を示す。2.2節では転移学習において転移元ドメインが有する知識を統計的に集約する手法として用いられる行列因子分解に関する関連研究を述べる。

### 2.1 転移学習に関する応用研究

特徴ベース・分離型の転移学習は協調フィルタリングに応用されている。協調フィルタリングは、ユーザに映画等のアイテムを推薦するための手法であり、ユーザのアイテムに対する評価値の行列（評価行列）を元に未評価のアイテムの予測評価値を求め、最終的にユーザにアイテムを推薦アイテムとしてランキング提示する手法である。協調フィルタリングにおける特徴ベース・分離型の転移学習[17][18]では、転移元ドメインにおいて得られるデータの種類、定義域または分布が転移目標ドメインと異なることを前提とし、転移元ドメインのデータを転移目標ドメインに適合するように変換し転移目標ドメインでの協調フィルタリングに利用する。[17][18]では、転移元ドメインで転移目標ドメインに有用な部分空間を求める際も、転移目標ドメインにおいて協調フィルタリングを行う際も、ともに行列因子分解が用いられる。

CST (Coordinate System Transfer)[17]は、転移元ドメインで得られた暗黙的な評価に関する知識を転移目標ドメインの明示的な評価に加える手法である。CSTでは、転移元ドメインの暗黙的な評価に関する評価行列からユーザ、アイテムそれぞれの因子を抽出し、転移目標ドメインでの行列因子分解において、転移元ドメインで得られた因子と行列因子分解により得られる因子が類似するように正則化項を加えることにより、暗黙的な評価の知識を反映させる。CBT (CodeBook Transfer)[18]は、転移元ドメインと転移目標ドメインが異なるアイテムカテゴリ（例えば、映画と本）に関する評価行列を有することを想定し、転移元ドメインの

アイテムカテゴリの評価に関する知識を転移目標ドメインに反映する。CBTでは、Tri-Factorizationを用い、転移元ドメインの評価行列からユーザグループとアイテムグループの関係を集約した因子の情報（コードブック）を作成し、転移目標ドメインに持ち込み行列因子分解に用いることで、転移元ドメインのアイテムカテゴリの評価の知識を転移目標ドメインに反映させる。

特徴ベース・統合型の転移学習も協調フィルタリングに適用されており、本論文の特徴ベース・分離型とはアプローチとは異なるが関連性があるため参照する。協調フィルタリングにおける特徴ベース・統合型の転移学習では、転移元ドメインのデータと転移目標ドメインのデータを収集して、同時に行列因子分解を行う。CMF (Collective Matrix Factorization) [19]は、転移元ドメイン、転移目標ドメインがそれぞれアイテムに関する補助データ、ユーザのアイテムに対する評価行列を有することを想定し、評価行列の行列因子分解において補助データから得られる知識を反映するものである。CMFでは、補助データの行列と評価行列を同時に行列因子分解する際、アイテムに関する因子の行列が同じになるように制約をかける。

上記の協調フィルタリングにおける応用事例において、以下の前提を置いている。

- CST[17]: ユーザ群とアイテム群が転移元ドメインと転移目標ドメインで一致
- CBT[18]: アイテム群が転移元ドメインと転移目標ドメインで排他
- CMF[19]: アイテム群が転移元ドメインと転移目標ドメインで一致、かつ、転移元ドメインと転移目標ドメインのデータが共同利用可能

本論文では転移元ドメインと転移目標ドメインの属性が部分一致している状況を想定しているが、上記手法はいずれも一致か排他の前提をとっておりそのまま適用することはできない。

## 2.2 行列因子分解に関する関連研究

転移元ドメインのデータをコンパクトに表現する部分空間を学習することは特徴ベースの転移学習の肝となるが、この問題は行列因子分解の問題として定式化できる。なかでも代表的な手法として非負値行列因子分解 (NMF: Nonnegative Matrix Factorization) [20]がある。NMFは、元の非負値行列 $X$ をふたつの低ランクの非負値行列 $U, V$ の積で近似する手法である。これにより、行列 $X$ を構成する各ベクトルを非負値の基底ベクトルの非負値係数による線形結合で表現することができる。

$$X \approx U^T V \quad s.t. U, V \geq 0$$

この行列因子分解は以下の非負値最小二乗問題として定式化できる。

$$\min_{U, V} \|X - U^T V\|^2 \quad s.t. U, V \geq 0$$

NMFは、行列の各ベクトルを基底ベクトルの線形結合で解釈できるという解釈の容易さという利点を持ち、さらに損失関数や正則化の工夫により基底ベクトルが元の行列に関する追加的な特徴を保存するように拡張することが可能である。

NMFを教師あり学習に適応させるため、データに付与された教師ラベルに関する追加的な特徴を保存するように拡張した手法が提案されている[21][22][24]。特にWangら[24]の手法は、Yangら[23]の提唱したグラフ埋め込みという学習フレームワークのひとつの応用事例にあたり、半教師あり学習が可能である特長を持つ。Yangら[23]は、グラフ埋め込みというラプラシアン行列の制約付き低ランク近似により、NMFを含む多くの部分空間の学習法に対する一般化を与えた。この手法は、Wangら[24]によって教師あり/半教師あり学習に拡張された。この拡張手法では、各データを頂点としてデータ間の類似度をエッジとするグラフ表現において同一の教師ラベルを有するデータ同士の類似度、異なる教師ラベルを有するデータ同士の非類似度等について制約を付与するように、元の損失関数に正則化項を加えている。本論文はWangら[24]の手法を採用し、詳細は次章で述べる。

## 3. 問題設定とアプローチ

本章では改めて本論文の問題設定を明確にし、本論文でのアプローチについて述べる。

### 3.1 問題設定

本論文で取り組む転移学習の問題設定は以下のようなものである。

互いに生データを連携することのできない2つのドメインを想定する。ドメインとはひとまとまりのデータが得られる領域のことである。ドメイン間の知識の転移方向の観点から、一方を元ドメイン、他方を目標ドメインと呼ぶこととする。以下のようなプロセスにより学習器 $f$ の改善を目指す。

1. 元ドメインにおいて元ドメインのデータ $D_S$ から統計処理 $g$ により知識 $K$ を得る
2. 知識 $K$ を元ドメインから目標ドメインに連携する
3. 知識 $K$ と目標ドメインのデータ $D_T$ を用いて学習器 $f$ を構築する

本論文では特に $D_S$ と $D_T$ に共通の属性がある場合を扱う。両ドメインの属性は一致しないものとし、より多くの属性を有するドメインを元ドメイン、より少ない属性を有するドメインを目標ドメインとする。簡単のため目標ドメインに固有の属性は無いものとする。また、学習器が解こうとするタスクのラベルデータを両ドメインにおいて用いることができるものとする。

元ドメインのデータ $D_S$ は属性データ $X_S$ とラベルデータ $Y_S$ からなる。 $X_S = (x_{S1}, x_{S2}, \dots, x_{Sn_S})^T \in \mathbb{R}_+^{n_S \times m_S}$  はサンプル

数 $n_S$ , 属性数 $m_S$ , 各要素が非負値であるような行列であるとする.  $X_S$ の一部には互いに排他的なクラスラベル  $Y_S = (y_{S1}, y_{S2}, \dots, y_{Sn_L})^T$  ( $n_L < n_S$ ) が付与されており, 簡単のため2クラスとする. すなわち,  $y_{Si} \in \{c_1, c_2\}, i = 1, 2, \dots, n_L$  である.

目標ドメインのデータ $D_T$ は属性データ $X_T$ とラベルデータ $Y_T$ からなる.  $X_T = (x_{T1}, x_{T2}, \dots, x_{Tn_T})^T \in \mathbb{R}_+^{n_T \times m_T}$ はサンプル数 $n_T$ , 属性数 $m_T$ , 各要素が非負値であるような行列であるとする. 簡単のため $X_T$ は全てにラベルが付与されているものとする. すなわち,  $Y_T = (y_{T1}, y_{T2}, \dots, y_{Tn_T})^T, y_{Ti} \in \{c_1, c_2\}$ である.

$X_S$ の属性の集合を $a(X_S)$ ,  $X_T$ の属性の集合を $a(X_T)$ とすると, 本論文の両ドメインの属性についての前提は,  $a(X_S) \supset a(X_T)$ と表せる.

### 3.2 NMF を用いた知識構築

本論文では上述の問題設定に対して, 知識 $K$ をNMFの因子 $V$ として構築する. 構築された因子 $V$ は, 属性データ $X_T$ とともに目標ドメインにおいて非負値最小二乗法(NNLS)を通じて未知属性値 $X'_T$ の推定に用いられる,  $X'_T$ は $D_T = \{X_T, Y_T\}$ と共に学習器 $f$ の構築に用いられる. 本論文におけるアプローチの概念図を1に示す.

元ドメインのラベルデータ活用を視野に入れるため, 因子 $V$ の構築には半教師ありNMFを用いる. [24]で示された定式化を本論文の問題設定に合わせて以下のように表せる.

$$\begin{aligned} & \min_{U, V} \|X - UV^T\|^2 + \\ & \alpha \sum_{i \neq j} \|U_i^1 - U_j^1\|^2 S_{ij} + \\ & \alpha \sum_{i \neq j} \|U_i^2 - U_j^2\|^2 S_{ij}^p + \\ & \beta \sum_{i \neq j} \|U_i^1 - U_j^1\|^2 S_{ij}^s + \\ & \alpha, \beta \geq 0 \\ & \text{s.t. } U, V \geq 0 \end{aligned}$$

第一項は一般的なNMFの目的関数であり, 元の行列 $X \in \mathbb{R}_+^{n \times m}$ を係数行列因子 $U \in \mathbb{R}_+^{n \times r}$ と基底行列因子 $V \in \mathbb{R}_+^{m \times r}$ の積で近似することを表す. 因子のランク $r$ は $r < \min(n, m)$ とし,  $U$ と $V$ ではそれぞれ特徴空間と事例空間の次元が圧縮された部分空間が得られる.  $X$ の行がユーザを表す場合, 得られる $V$ はユーザの次元が圧縮されているという観点で統計処理されている. 元ドメインにおいて実施するため,  $X = X_S$ である.

第二~第四項はいずれも部分空間におけるデータ点の距離を操作する正則化項である.

第二, 第三項はラベル情報を部分空間に反映させる正則化項である. 第二項はクラス内緊密性を, 第三項はクラス間分離性を与える. 教師情報の反映に際して, 係数行列因

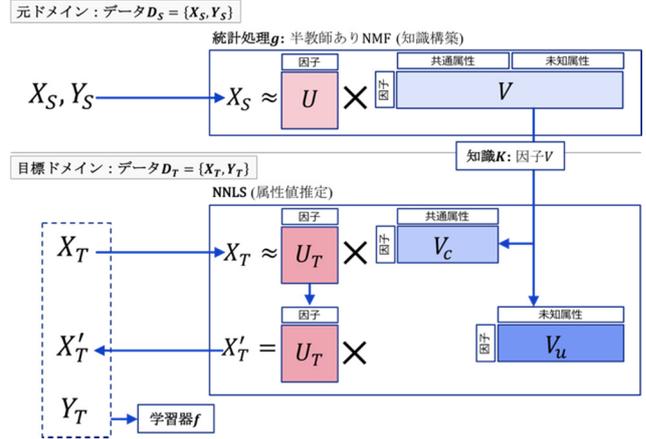


図1: 本論文におけるアプローチの概念図

子 $U$ の部分空間を $U = [U^1, U^2], U^1 \in \mathbb{R}_+^{n \times q}, U^2 \in \mathbb{R}_+^{n \times (r-q)}$ と分割し,  $U^1$ をラベル情報の反映に用いる.

第二項では,  $U^1$ の空間において同一クラスのデータ点を近づけることでクラス内緊密性を高める.  $S_{ij}$ は,

$$S_{ij} = \begin{cases} 1 & \text{if } i \in N_{k1}(j) \text{ or } j \in N_{k1}(i) \\ 0 & \text{otherwise} \end{cases}$$

ここで,  $N_{k1}(i)$ はサンプル $i$ と同一のクラス内における $k1$ -最近傍点の添字集合である.

第三項では, 補助的な $U^2$ の空間において異なるクラスのデータ点を近づけることでクラス間分離性を代替する. これは, 本来異なるクラスのデータ点は $U^1$ の空間において遠いことが望ましく, それは  $\max_{i \neq j} \|U_i^1 - U_j^1\|^2 S_{ij}^p$ と表現できるが, 最小化問題として定式化するためである.  $S_{ij}^p$ は,

$$S_{ij}^p = \begin{cases} 1 & \text{if } (i, j) \in P_{k2}(c_1, c_2) \\ 0 & \text{otherwise} \end{cases}$$

ここで,  $P_{k2}(c_1, c_2)$ は異なるクラス $c_1 - c_2$ 間における $k2$ -最近接ペアの添字ペア集合である.

第四項はラベル情報によらず元の空間における近傍関係を反映する正則化項である.  $S_{ij}^s$ は,

$$S_{ij}^s = \begin{cases} 1 & \text{if } i \in N_{k3}(j) \text{ or } j \in N_{k3}(i) \\ 0 & \text{otherwise} \end{cases}$$

ここで,  $N_{k3}(i)$ はサンプル $i$ のクラスに依らない $k3$ -最近傍点の添字集合である.

なお, 目的関数の第四項を除外して $X$ をラベルの付与されたものに限定, すなわち $X = (x_{S1}, x_{S2}, \dots, x_{Sn_L})$ とすれば教師ありの手法となり, 目的関数の第一項のみを残せば教師なしの手法となる.

目的関数を非負値制約の元で最適化する手法については, [25]の乗法更新則や, 勾配法の加法更新則において勾配にランブ関数(ReLU)を適用するものがあるが[26], 本論文では数値計算フレームワークTensorFlow[27]を活用するため, [28]のパラメータ化 $U = E^2, E \in \mathbb{R}_+^{n \times r}, V = F^2, F \in \mathbb{R}_+^{m \times r}$ により非負性を与え, 勾配法の一つであるAdam[29]を用いることとした.

### 3.3 知識を用いた属性値推定

3.2 の方法によって元ドメインで構築した知識としての因子 $V$ は、属性 $a(X_S)$ の出現パターンの基底と見ることができる。これら基底の非負値係数による線形結合が生データを表すよう構築されたものが因子 $V$ であるため、目標ドメインにおいても非負値係数を求めることで、未知属性 $a(X_S) \setminus a(X_T)$ の推定値 $X'_T \in \mathbb{R}_+^{n_T \times m_S - m_T}$ を得ることができる。以下にその方法を示す。

まず、因子 $V$ を共通属性と未知属性の部分に分ける。すなわち、

$$\begin{aligned} V^T &= (V_c^T, V_u^T) \\ &s. t. \\ a(V_c) &= a(X_S) \cap a(X_T) \\ a(V_u) &= a(X_S) \setminus a(X_T) \\ V_c &\in \mathbb{R}_+^{m_T \times r} \\ V_u &\in \mathbb{R}_+^{m_S - m_T \times r} \end{aligned}$$

次に、因子の共通属性部分 $V_c$ が $X_T$ を最も良く表すよう非負値係数行列 $U_T$ を非負値最小二乗法 (NNLS) により求める。すなわち、

$$\min_{U_T} \|X_T - U_T V_c^T\|^2, U_T \geq 0$$

最後に、非負値係数行列 $U_T$ と因子 $V$ の未知属性部分 $V_u$ との積により未知属性の推定値を得る。すなわち、

$$X'_T = U_T V_u^T$$

目標ドメインにおける学習器 $f$ の改善は、 $X_T$ だけでなく $X'_T$ を用いて特徴量を作成することによって実現する。

### 3.4 ロバストな損失関数

本論文では、両ドメインのデータとして属性の分布がロングテールとなる Web アクセスログを想定し、知識構築の NMF と属性値推定の NNLS においてロバストな損失関数を用いる。

#### 3.4.1 二乗損失 (L2 損失)

二乗損失 (L2 損失) 関数は誤差 $e_{ij}$ を用いて以下のように表せる。

$$J_{L2}(e_{ij}) = \sum_{i,j} e_{ij}^2$$

フロベニウスノルムは二乗損失関数を用いていることに相当する。誤差を二乗するため、外れ値の影響を大きく受ける。

#### 3.4.2 絶対損失 (L1 損失)

絶対損失 (L1 損失) 関数は以下のように表せる。

$$J_{L1}(e_{ij}) = \sum_{i,j} |e_{ij}|$$

誤差の絶対値に着目するため、二乗関数ほど外れ値の影響を受けにくい特徴を有する。

#### 3.4.3 Huber 損失

Huber 損失関数は以下のように表せる。

$$\begin{aligned} J_{Huber}(e_{ij}) &= \sum_{i,j} H(e_{ij}) \\ H(x) &= \begin{cases} \frac{1}{2}x^2 & \text{if } |x| < \delta \\ \delta \left( |x| - \frac{\delta}{2} \right) & \text{otherwise} \end{cases} \end{aligned}$$

絶対損失では誤差 0 で微分不可能なため、勾配法で解を安定させるには学習率を減衰させる必要があるが、Huber 損失では $\delta$ 未満の誤差に対して二乗損失として機能するため解を安定させやすい。また、Huber 損失は $\delta$ 以上の誤差に対しては絶対損失として機能するため外れ値に対する耐性もある。

## 4. 実験と考察

本節では、半教師あり NMF を用いた転移学習を Web アクセスログに基づく引越予測タスクに適用し、教師ラベルの効果と損失関数の効果について調べた実験について述べる。

### 4.1 データ

データは広告配信プラットフォームの Web アクセスログ 3 ヶ月分と、その翌月のイベントフラグを用いた。Web アクセスログはユーザごとの URL アクセスの履歴であり、URL は恣意的に幾つかの URL カテゴリに分類されている。Web アクセスログのユーザ数は 14,257 人、URL カテゴリは 1,025 個である。イベントフラグは各ユーザのイベント有無を示す二値であり、Web アクセスログの一部のユーザにのみ付与されている。イベントフラグはアンケートで引越の有無について回答を得たもので、ユーザ数は 4,092 である。Web アクセスログのうちアンケート回答を得たユーザを除く 9,819 人はサンプリングされたものである。

### 4.2 前処理

Web アクセスログはユーザと URL カテゴリをそれぞれ文書と単語と見なした Bag of Words 表現で行列化し、行列の各要素は 3 ヶ月間のアクセス数の合計とした。この Web アクセス数行列を以下の要領で分割することにより異なるドメインのデータを模擬した。まず、イベントフラグの付与されたユーザを元ドメインユーザと目標ドメインユーザに無作為に二分した。次に、イベントフラグの付与されていないユーザを全て元ドメインユーザとした。最後に、URL カテゴリを共通属性と未知属性に無作為に二分し、目標ドメインユーザの未知属性のアクセス数は推定対象とした。この処理により、 $X_S$ は 12,000×1,025、 $X_T$ は 2,257×512、 $X'_T$ は 2,257×513 となった。また、 $X_S$ をラベルありユーザに絞った $X_L$ は 2,181×1,025 となった。

### 4.3 評価対象

知識構築手法としてラベル情報の利用形態に応じて以下の 4 通りの NMF を評価した。

1. uNMF\_L:  $X_L$ を入力とした教師なし NMF

2. uNMF\_S:  $X_S$ を入力とした教師なし NMF
3. sNMF:  $X_L$ と $Y_L$ を入力とした教師あり NMF
4. ssNMF:  $X_S$ と $Y_L$ を入力とした半教師あり NMF

また、知識構築および属性値推定における損失関数として以下の3通りを評価した。

- a. L1: 絶対損失
- b. L2: 二乗損失
- c. Huber: Huber 損失

本評価では上記の組合せである12通りを評価した。

#### 4.4 評価方法

4.3の手法で知識構築と属性値推定を行い、推定された未知属性値と目標ドメインのデータを用いて学習器を構築し、汎化性能を調べた。知識構築のパラメータは、因子数  $r = \{50, 100, 150\}$ ,  $q = r/2$ ,  $k_1 = 179$ ,  $k_2 = 60$ ,  $k_3 = 2181$ ,  $\alpha = 1$ ,  $\beta = 1$ とした。 $k_1 \sim k_3$ は[23]を参考に経験的に定めた。別途L2正則化項を係数3で追加した。Huber損失では $\delta = 1$ とした。知識構築と属性値推定において変数の初期化は平均0、標準偏差1の乱数により行い、最適化はAdamのパラメータを学習率0.1, beta1=0.9, beta2=0.999, epsilon=1e-8とし、更新回数の上限を100回とし、目的関数の改善が前回更新時の0.1%以下となることが10回を超えた場合は打ち切りとした。特徴量は各URLカテゴリへのアクセス有無を示す0/1とし、得られた $X_T'$ の分布がゼロに偏っている場合は非ゼロ要素ならば1とし、分布がゼロに偏っていない場合は1以上なら1とした。学習器にはXGBoost[30]を用いた。汎化性能の精度指標としては3-fold cross validationを100回実施しF1-score, AUC (area under the ROC curve: AUC)それぞれの平均値を算出した。F1-scoreの算出に際しては予測値0.5以上をPositiveとした。

#### 4.5 結果と考察

知識構築手法と損失関数の組合せに応じた学習器の性能として、F1-scoreの比較結果を表1に、AUCの比較結果を表2に示す。損失関数を固定したときに精度指標が最も良くなるものを太字で示し、知識構築手法を固定したときに精度指標が最も良くなるものを下線で示した。参考までに、元ドメインの知識を用いず目標ドメインのデータ  $D_T = \{X_T, Y_T\}$ のみを用いて評価した学習器の精度指標は、F1-score = 0.0854, AUC = 0.6392であった。

ロバストな損失関数を用いる効果について、AUCの下線部に着目すると、どの知識構築手法においてもHuber損失を用いるケースが最良となった。このような結果となった要因としては、データに対して想定していたロングテールという特徴が当てはまったことが考えられる。

教師あり・半教師あり手法を通じてラベル情報を用いる効果について、AUCの太字に着目すると、損失関数がL2またはL1の場合は元ドメインの属性データをすべて用いた教師なしNMFが最良となったが、損失関数がHuberの

表1: F1-scoreの比較

知識構築手法	r	L2	L1	Huber
sNMF	50	0.0753	0.0493	0.0971
	100	0.0729	0.0638	0.0964
	150	<u>0.0979</u>	0.0691	0.0878
ssNMF	50	0.0858	0.0653	0.1046
	100	0.0811	0.0643	<b>0.1066</b>
	150	<b>0.1021</b>	0.0702	0.0911
uNMF_L	50	0.0634	0.0473	<u>0.1026</u>
	100	0.0769	0.0679	0.1012
	150	0.0971	0.0703	0.0968
uNMF_S	50	0.0947	0.0620	<u>0.1044</u>
	100	0.0789	0.0620	0.0871
	150	0.0768	<b>0.0837</b>	0.0871

表2: AUCの比較

知識構築手法	r	L2	L1	Huber
sNMF	50	0.6007	0.5845	<u>0.6261</u>
	100	0.5919	0.5668	0.6162
	150	0.5905	0.6014	0.6217
ssNMF	50	0.5968	0.6131	0.6225
	100	0.6026	0.6060	<b>0.6404</b>
	150	0.6131	0.6105	0.6109
uNMF_L	50	0.5705	0.5788	0.6253
	100	0.5873	0.6090	0.6265
	150	0.5861	0.5885	<u>0.6312</u>
uNMF_S	50	0.6144	<b>0.6205</b>	<u>0.6273</u>
	100	0.5996	0.5893	0.6143
	150	<b>0.6245</b>	0.6200	0.6073

場合は半教師ありNMFが最良となった。このような結果となった要因としては、目的関数のどの項が支配的であったかが影響していると考えられる。したがって、ラベル情報の効果を引き出すための条件を特定するためにはハイパーパラメータチューニングを通じたさらなる評価が必要だと考える。

F1-scoreに着目すると、ssNMF(r=100, Huber)は元ドメインの知識を用いない場合と比べて1.25倍となっており、応用的な観点から有益な精度向上が得られた。

#### 5. おわりに

本論文ではドメイン間のデータの属性が部分一致しているケースにおいてNMFを用いた転移学習による目標ドメインの二項分類タスクの精度向上について調べた。1つのアプローチとして、元ドメインから目標ドメインに連携する知識としてNMFの因子を用い、目標ドメインにおいてNNLSを用いて未知属性値を推定し、学習器の特徴量とする方法を示した。知識構築において半教師あり手法を通じてラベル情報を用いることの効果と、知識構築と属性値推定においてロバストな損失関数を用いることの効果の評価した。その結果、半教師あり手法とHuber損失の組み合わせが最善となり、転移学習を用いない場合と比較してF1-scoreが1.25倍となることが確認された。

## 参考文献

- [1] “Apache kafka”. <https://kafka.apache.org/>.
- [2] “Fluentd”. <http://www.fluentd.org/>.
- [3] “HDFS”. [https://hadoop.apache.org/docs/r1.2.1/hdfs\\_design.html](https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html)
- [4] “Apache Hadoop”. <http://hadoop.apache.org/>.
- [5] “Apache Spark”. <https://spark.apache.org/>.
- [6] “Apache Spark MLlib”. <https://spark.apache.org/mllib/>.
- [7] “Theano”. <http://deeplearning.net/software/theano/>.
- [8] “TensorFlow”. <https://www.tensorflow.org/>.
- [9] Linden, G., Smith, B. and York, J.. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 2003, vol. 7, no. 1, p. 76-80.
- [10] Bennett, J. and Lanning, S.. The netflix prize. *KDD cup and workshop*, vol. 2007, p. 35.
- [11] 小野智弘, 黒川茂莉, 本村陽一, 麻生英樹. ユーザ嗜好の個人差と状況差を考慮した映画推薦システムの実現と評価. *情報処理学会論文誌*, 2008, vol. 49, no. 1, p. 130-140.
- [12] Gunosy. <https://gunosy.com/>.
- [13] 和田健太郎, 小林桂子, 桑原雅夫. プローブ車両軌跡データのみを用いた交通の量的把握. *生産研究*, 2015, vol. 67, no. 2, p. 143-147.
- [14] Vapnik, V. N. and Chervonenkis, A. Y.. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, 1971, vol. 16, no. 2, p. 264-280.
- [15] Schein, A. I., Popescul, A., Ungar, L. H. and Pennock, D. M.. Methods and metrics for cold-start recommendations. *ACM SIGIR*, 2002, p. 253-260.
- [16] 神尾 敏弘. 転移学習. *人工知能学会誌*, 2010, vol. 25, no. 4, p. 572-580.
- [17] Pan, W., Xiang, E. W., Liu, N. N. and Yang, Q.. Transfer learning in collaborative filtering for sparsity reduction. *AAAI*, 2010, vol. 10, p. 230-235.
- [18] Li, B., Yang, Q. and Xue, X.. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. *IJCAI*, 2009, p. 2052-2057.
- [19] Singh, A. P., and Gordon, G. J.. Relational learning via collective matrix factorization. *ACM SIGKDD*, 2008, p. 650-658.
- [20] Lee, D. D. and Seung, H. S.. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature*, 1999, vol. 401, no. 6755, p. 788-791.
- [21] Jia, Y. W. Y. and Turk, C. H. M.. Fisher Non-Negative Matrix Factorization for Learning Local Features. *Asian Conf. Comp Vision*, 2004, p. 27-30.
- [22] Kotsia, I., Zafeiriou, S. and Pitas, I.. A Novel Discriminant Non-Negative Matrix Factorization Algorithm with Applications to Facial Image Characterization Problems. *IEEE Trans. Information Forensics and Security*, 2007, vol. 2, no. 3, p. 588-595.
- [23] Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q. and Lin, S.. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no. 1, p. 40-51.
- [24] Wang, C., Yan, S., Zhang, L. and Zhang, H.. Non-negative semi-supervised learning. In *Artificial Intelligence and Statistics*, 2009, p. 575-582.
- [25] Lee, D. D. and Seung, H. S.. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 2001, p. 556-562.
- [26] Lin, C. J.. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 2007, vol. 19, no. 10, p. 2756-2779.
- [27] Abadi, M. et al.. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. 2016, <http://tensorflow.org/>.
- [28] Chu, M., Diele, F., Plemmons, R. and Ragni, S. Optimality, computation, and interpretation of nonnegative matrix factorizations. *SIAM Journal on Matrix Analysis*, 2004.
- [29] Kingma, D., and Ba, J.. Adam: A method for stochastic optimization. *ICLR*, 2015.
- [30] Chen, T., and Guestrin, C.. Xgboost: A scalable tree boosting system. *ACM SIGKDD*, 2016, p. 785-794.