

# 品質の異なる二つのデータ集合間の転移学習の解析

## An Analysis of Transfer Learning between a Pair of Datasets with Different Qualities

赤穂 昭太郎\*  
Shotaro Akaho

神島 敏弘\*  
Toshihiro Kamishima

**Abstract:** It is often the case that we have a pair of datasets, where one consists of few high-quality data and the other consists of many low-quality data. Kamishima et al. has proposed a bagging-based learning algorithm to synthesize such a pair of data sets. However, it has not been clear when the learning algorithm improves the estimation obtained by using only the few high-quality dataset. We analyze a simple exponential family model, and prove that the bagging-based learning algorithm does not work for this model, but an appropriate modification improves the estimation drastically.

**Keywords:** transfer learning, asymptotic analysis, statistical estimation, exponential family

### 1 まえがき

近年、インターネットの発達などによって大規模なデータを収集することは比較的容易である。例えば、映画の好みなどユーザの嗜好を集めたデータや、スーパーマーケットでの消費者の購買履歴といったものである。しかしながら、通常そうして集められたデータにはさまざまな種類のユーザに対する嗜好が混在しており、また十分な管理下で集められたものでないためデータの質も低いことが多い。そのようなデータから、例えば各ユーザごとにきめ細かく商品の推薦などを行おうとしても、十分な精度が得られないことがある。

一方、これらのデータとは別に、ある特定の種類のユーザ群に対して十分管理された状況下でデータを収集することを考えよう。ただし、これには大きなコストがかかるため、それほど大きなサイズのデータを集めることはできず、統計的な推定精度は低くなってしまふ。

そこで、もともとの大規模なデータを補助情報として用いることによって、この新たに集めた質の高いデータに対する統計的な推定精度を向上できないかという問題が考えられる。神島ら [1] によって提案された「飼いならし学習」はそのような問題意識から考え出された枠組

みであり、学習分野で古くから研究されてきた転移学習 [6] の特別な場合として位置づけることもできる。

また、ユーザの嗜好や購買といった問題に限らず、自然科学実験などにおいても「コストが高くて収集に時間のかかる品質の高い少数データ集合」と「コストが安くて収集が容易だが品質の低い大規模データ集合」があるという状況は頻繁に発生する。

神島らは、バギングを応用したアルゴリズムを用いて、実データを使ったタグ付け問題に対して有効性を示したが [1]、理論的にはどのような場合にうまくいくかが明らかではなかった。そこで本稿では、この問題の非常に単純な数理モデルに対して解析を行い、神島らによって提案されたアルゴリズムがうまくいくためには、少なくともある種の非線形性が必要なことと、それが満たされない場合にはそれを改良したアルゴリズムによって精度のよい推定ができることを示す。

### 2 品質の異なるデータ集合のモデル化

管理された状況下で収集されるデータセットは、確率変数  $x$  に対するある確率分布（ターゲット分布） $p_T(x)$  に従って生成された独立なサンプル集合  $T = (x_1, x_2, \dots, x_{|T|})$  とする。一方、大量に集められたデータの中には  $p_T(x)$  に従うサンプルと、そうでない確率分布（非

\*産業技術総合研究所脳神経情報研究部門, 305-8568 茨城県つくば市梅園 1-1-1 中央第 2, tel. 029-861-5549, e-mail s.akaho@aist.go.jp, mail@kamishima.net

The National Institute of Advanced Industry and Scientific Technology, Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568

ターゲット分布)  $p_N(x)$  に従うサンプルが混在しているとす。すなわち,

$$p_S(x) = \alpha p_T(x) + (1 - \alpha)p_N(x) \quad (1)$$

という混合分布に従って独立に生成されたサンプル  $S = (x'_1, x'_2, \dots, x'_{|S|})$  であるとする。具体的にどのような分布を考えるかについては、後の節で詳細化する。

一般に  $T$  のサンプル数  $|T|$  は、 $S$  のサンプル数  $|S|$  に比べてずっと小さい。また、多くの状況では  $\alpha$  は 1 に比べてかなり小さい。さらに、確率変数  $x$  は、推薦システムやタグ付けなど多くの例題では入力と出力のペアからなる教師あり学習の場合が多いが、本稿では入力と出力の区別は特にせず解析を行い、教師あり学習については最後に違いなどについて論じる。

さて、以上のようなモデル仮定のもとで、やりたいことは、二つのデータ集合  $S, T$  に基づいてターゲット分布  $p_T(x)$  に関する精度の高い推測を行うことである。

神鷹らは、バギングに基づいたアルゴリズムを提案し、タグ付け問題に対して有効性を示した。ここでは、上記の定式化に合わせて単純化した形でそのアルゴリズムを示す。

### TrBagg アルゴリズム (簡略版)

1. データ集合  $T$  から確率分布を推定し、 $\hat{p}_T(x)$  と置く。
2. データ集合  $S$  からブートストラップサンプル集合  $B$  を生成する。
3.  $B$  から確率分布を推定し、 $\hat{p}_B(x)$  と置く。
4.  $\hat{p}_T(x)$  と  $\hat{p}_B(x)$  の距離を測り、あるしきい値以下であれば  $\hat{p}_B(x)$  を受取する。
5. 上記ステップ 2 ~ 4 を繰り返し、受取された確率分布を統合して最終的な確率分布の推定値  $\hat{p}_{\text{TrBagg}}(x)$  を得る。

上記アルゴリズムの最後のステップにおける「統合」の意味は、後の節でモデルをより詳細化する際に明確にする<sup>1</sup>。

さて、もし最終的に得られた  $\hat{p}_{\text{TrBagg}}(x)$  が、 $T$  だけを使って推定された  $\hat{p}_T(x)$  よりもデータを生成した分布  $p_T(x)$  に近ければ  $S$  を使った効果があったということになる。

<sup>1</sup>なお、オリジナルの TrBagg アルゴリズムでは  $B$  から作るのは識別器であり、最終的な統合は識別器の多数決によって得られる。また、このほかにも性能を向上させるためのいくつかのヒューリスティクスが導入されている。

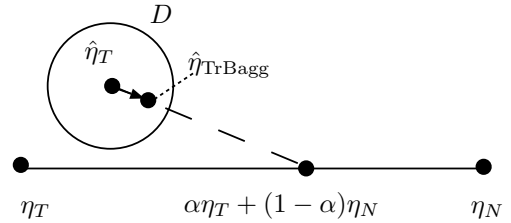


図 1: 指数分布族に対する TrBagg アルゴリズムの振る舞い

## 3 指数分布族の場合

ここでは、データを生成する確率分布がある指数分布族の要素である場合には、TrBagg アルゴリズムがそのままではうまくいかない可能性が高いことを示す。

命題 1  $p_T(x), p_N(x)$  の両方が、ある指数分布族の要素であり、 $\alpha$  が 1 より十分小さいとき、最尤推定に基づいた TrBagg アルゴリズムによって  $\hat{p}_T(x)$  を改良できる確率は約  $1/2$  である。つまり、改良できるかできないかが五分五分である。

上記の性質は漸近論的に理解できる。以下にそのあらましを述べよう (図 1 参照)。指数分布族は

$$p(x; \theta) = \exp(\theta \cdot r(x) - \psi(\theta) + C(x)) \quad (2)$$

という分布形を持つ確率分布である。

まず、指数分布族の確率分布は、期待値パラメータ  $\eta = E_{p(x; \theta)}[r(x)]$  を座標系とする空間上の 1 点として表現できることに注意する。すると、 $p_T(x)$  および  $p_N(x)$  もそれぞれ空間上の点  $\eta_T, \eta_N$  で表現される。

同様に、一つ一つのサンプル  $x$  に対しても  $r(x)$  を  $\eta$  座標に対応させることにより空間上の点とみなすことができる。いくつかのサンプル点があったとき、パラメータの最尤推定はそれらの  $\eta$  座標  $r(x)$  の重心 (平均) で与えられる。

さて、我々の問題では、まずデータ集合  $T$  から最尤推定により  $\eta_T$  を推定するわけだが、その推定値  $\hat{\eta}_T$  は、漸近的には  $\eta_T$  の周りに正規分布をすることが知られている。

一方、データ集合  $S$  から得られたブートストラップ集合  $B$  の中には  $p_T(x)$  に由来するもの  $B_T$  と  $p_N(x)$  に由来する  $B_N$  の二つがある。 $B$  の要素数を  $|B|$  とし、そのうち  $k$  個が  $p_T(x)$  由来のものだとする。 $k$  は二項分布に従う確率変数として与えられるが、近似的に  $k = |B|\alpha$  というピーク付近にあるとして固定して考える<sup>2</sup>。すると、 $B_T$  の  $\eta$  座標重心  $\eta_{B_T}$  は漸近的に  $\eta_T$  の

<sup>2</sup>この部分を厳密化しても議論は定性的には影響を受けない

まわりに正規分布し、 $B_N$  の  $\eta$  座標重心  $\eta_{B_N}$  は漸近的に  $\eta_N$  のまわりに正規分布する。そして、 $B$  全体の  $\eta$  座標重心は正規分布の和の分布なので、やはり正規分布となり、その中心を  $\eta_B$  とすると、 $k = |B|\alpha$  の仮定の下で

$$\eta_B = \alpha\eta_T + (1 - \alpha)\eta_N, \quad (3)$$

すなわち  $p_T(x)$  と  $p_N(x)$  のデータ集合  $S$  上での比率に応じた重心座標として与えられる。

以上の議論はすべてのブートストラップ集合の推定量の分布であるが、TrBagg アルゴリズムでは、 $p_T(x)$  のデータ集合  $T$  に基づいた推定量  $\hat{\eta}_T$  から一定距離にある推定量だけを取り出して統合する。ここで、「統合する」というのを、推定量の重心を取ることと定義する。 $\hat{\eta}_T$  から一定距離の範囲を  $D$  と書くと、推定量だけの重心は、 $\eta_B$  を中心とする正規分布  $N[\eta_B]$  の  $D$  だけに限定した期待値

$$\hat{\eta}_{\text{TrBagg}} = \frac{\int_D \eta N[\eta_B] d\eta}{\int_D N[\eta_B] d\eta} \quad (4)$$

であり、これは正規分布の形状から、 $\hat{\eta}_T$  より  $\eta_B$  の方向にシフトしたものになることがわかる。従って、 $\hat{\eta}_T$  から見て  $\eta_B$  の方向と  $\eta_T$  の方向とが同じ方向を向いていることが推定量を改善するために必要である。

さて、 $\eta_B$  は (3) からわかるように、 $\eta_T$  と  $\eta_N$  が十分離れており、 $\alpha$  が 1 より十分小さいとき、 $T$  だけを使った推定量  $\hat{\eta}_T$  よりも  $\eta_T$  から離れたところにある。このときは、 $\hat{\eta}_T$  は  $\eta_T$  を中心とする正規分布に従うので、 $\hat{\eta}_{\text{TrBagg}}$  が  $\hat{\eta}_T$  に比べて  $\eta_T$  に近くなるかどうかは五分五分である。

## 4 正規分布の位置母数の推定の解析

### 4.1 指数分布族に対する改良アルゴリズム

前節での考察から、指数分布族のようにパラメータがデータの統計量の線形和で表されるような単純な場合には TrBagg 法はうまくいかないことがわかった。その主な原因は、 $p_N(x)$  に由来するサンプルがある程度含まれていても、サンプル平均がたまたま  $\hat{p}_T(x)$  の近くに分布する可能性は高いということである。

そこで本稿では、指数分布族のように各データ点もパラメータ空間上の点として表される場合に以下のようなアルゴリズムを考える。

#### データフィルタリングアルゴリズム

1. データ集合  $T$  から確率分布を推定し、その  $\eta$  座標を  $\hat{\eta}_T$  と置く。

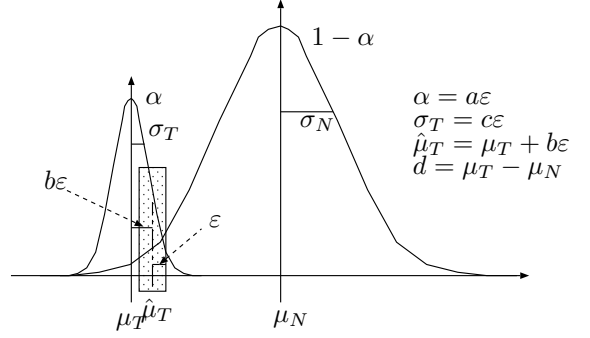


図 2: 正規分布の位置母数推定

2. データ集合  $S$  のサンプル点の座標表現  $r(x)$  のうち、 $\hat{\eta}_T$  との距離があるしきい値以下のものだけを選び出す。
3. 選び出されたサンプル点の  $r(x)$  の平均を取って  $\hat{\eta}_{\text{Filter}}$  とし、これをパラメータとする確率分布を  $p_T(x)$  の推定値とする。

このアルゴリズムのポイントは、しきい値以下となるような  $r(x)$  だけに限定することによって  $p_T(x)$  から遠く離れたサンプルを排除するようにした点である。ただし、実際にはしきい値以下のサンプル数が十分確保できない場合も考えられるため、距離に応じた重み付き期待値をとるなどのバリエーションが考えられる。

### 4.2 正規分布の位置母数推定の解析

上で述べたアルゴリズムが推定量を改良できるかどうかを見るために、簡単なモデルを使った解析を行う (図 2)。

まず、ターゲット分布  $p_T(x)$  は正規分布  $\mathcal{N}[\mu_T, \sigma_T^2]$  とし、非ターゲット分布  $p_N(x)$  は別の正規分布  $\mathcal{N}[\mu_N, \sigma_N^2]$  とする。

今、ターゲット分布の位置パラメータ  $\mu_T$  を求めることを目標としよう。管理されたデータ集合  $T$  から推定されたパラメータの値を  $\hat{\mu}_T$  とする。データフィルタリングアルゴリズムでは、データ集合  $S$  のサンプルのうち  $\hat{\mu}_T$  との距離があるしきい値  $\epsilon$  以下のものを選び出し、その期待値  $\hat{\mu}_{\text{Filter}}$  を  $\mu_T$  の推定値とする。

ここで、データ集合  $S$  は十分大きいものと (つまり、無限個のサンプルが得られると) 仮定する。

以下では、 $\epsilon$  は微量量として、 $\epsilon$  に関して漸近的な  $\hat{\mu}_{\text{Filter}}$  の振る舞いを調べるが、しきい値以外の量も微量量として扱うのが適当と思われるものは  $\epsilon$  で表現する。まず、 $\hat{\mu}_T$  は  $\mu_T$  のまわりで分布すると考えられるので、確率変数  $b$  を取り、

$$\hat{\mu}_T = \mu_T + b\epsilon \quad (5)$$

と置く．また，管理されたデータではばらつきが比較的小さいと思われるため，

$$\sigma_T = c\varepsilon \quad (6)$$

のようにターゲット分布の標準偏差が  $\varepsilon$  オーダーであることを仮定する．なお， $\hat{\mu}_T$  の  $\mu_T$  からのずれを表す  $b\varepsilon$  は  $\sigma_T$  の大きさと  $T$  のサンプル数  $|T|$  に依存しており， $b$  は漸近的に平均 0 の正規分布に従い，

$$E[b^2] \simeq E\left[\frac{c^2}{|T|}\right] \quad (7)$$

が成立するので，比較的小さい  $|T|$  でも一般に  $b < c$  となるのが普通である．データ集合  $S$  に関しては，ターゲット分布に由来するデータ集合の割合を表す  $\alpha$  も小さいことが多いので

$$\alpha = a\varepsilon \quad (8)$$

と置く．

$S$  のサンプル数が無限という仮定のもとでは，サンプル平均は期待値で置き換えられるため，

$$\hat{\mu}_{\text{Filter}} = \int_{\hat{\mu}_T - \varepsilon}^{\hat{\mu}_T + \varepsilon} xp_S(x) dx \quad (9)$$

が求める推定量となる．これを  $\varepsilon$  について漸近展開したときに，

$$\hat{\mu}_{\text{Filter}} = \mu_T + \gamma b\varepsilon + O(\varepsilon^2) \quad (10)$$

となって， $|\gamma| < 1$  であれば  $\hat{\mu}_{\text{Filter}}$  は  $\hat{\mu}_T$  よりもよい推定量であると言える．

実際に  $\gamma$  を計算すると，

$$\gamma = \frac{\phi(d, \sigma_N^2) + \frac{c^2 a}{4b} \{\phi(b-1, c^2) - \phi(b+1, c^2)\}}{\phi(d, \sigma_N^2) + \frac{a}{4} \{\text{erf}(\frac{\sqrt{2}(b+1)}{2c}) - \text{erf}(\frac{\sqrt{2}(b-1)}{2c})\}} \quad (11)$$

となる．ただし，ここで  $\phi(\mu, v)$  は平均  $\mu$ ，分散  $v$  の原点における密度の値

$$\phi(\mu, v) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{\mu^2}{2v}\right) \quad (12)$$

を表し， $\text{erf}$  は誤差関数

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \quad (13)$$

を表す．また， $d$  は二つの正規分布の平均の差

$$d = \mu_N - \mu_T \quad (14)$$

とする．

上記の式は性質が分かりづらいので  $b$  が 1 より十分小さいと仮定して  $b$  で Taylor 展開すると，

$$\gamma = \frac{\phi(d, \sigma_N^2) + a\phi(1, c^2)}{\phi(d, \sigma_N^2) + \frac{a}{2}\text{erf}(\frac{\sqrt{2}}{2c})} + O(b^2) \quad (15)$$

となり， $b$  の 1 次項は 0 になっている．

さらに， $d$  が  $\sigma_N$  に比べて十分大きいとし（例えば 3 倍以上）， $c$  が十分小さいとすると（例えば 1/3 程度以下），分母の第 1 項はほぼ 0 となり，第 2 項は  $a/2$  にほぼ等しいので，近似的に

$$\gamma \simeq 2 \frac{\phi(d, \sigma_N^2)}{a} + 2\phi(1, c^2) \quad (16)$$

という式になる．以上をまとめると以下ようになる．

命題 2 二つの分布の平均の距離が  $\sigma_N$  に比べて十分大きく， $p_T(x)$  の分散の大きさを調節するパラメータ  $c$  が十分小さければ  $\gamma \ll 1$  となり，データフィルタリングアルゴリズムによって得られた推定量が  $\hat{\mu}_T$  を改善できる．

おそらく，TrBagg がタグ付け問題に対してうまく働いたのも，識別問題などでは TrBagg アルゴリズムがデータフィルタリングアルゴリズムと同じような効果 ( $p_N(x)$  由来の外れたデータが含まれていると識別器が受理されない) があつた可能性があるからと考えられる．

### 4.3 数値実験

上記のアルゴリズムが実際にどれくらい解を改善できるかを調べるために人工データを使った数値実験を行った．

推定量  $\mu$  が得られたとき，それが  $\hat{\mu}_T$  をどれだけ改善したかを見るのに

$$R(\mu) = \log \frac{|\mu - \mu_T|}{|\hat{\mu}_T - \mu_T|} \quad (17)$$

という指標を用いる．これは 0 のとき  $\hat{\mu}_T$  と同じ，正では悪く，負になると改善されていることを表す．

以下で，いろいろパラメータを変動させる実験を行うが，変動させないパラメータのデフォルト値は以下のように定める．

- $T$  のサンプル数 10
- $S$  のサンプル数 1000 (このうち  $p_T(x)$  より発生させたもののサンプル数 100,  $p_N(x)$  より発生させたもののサンプル数 900)
- $\mu_T = 0.0$ ,  $\sigma_T^2 = 0.1^2$
- $\mu_N = 1.0$ ,  $\sigma_N^2 = 0.5^2$

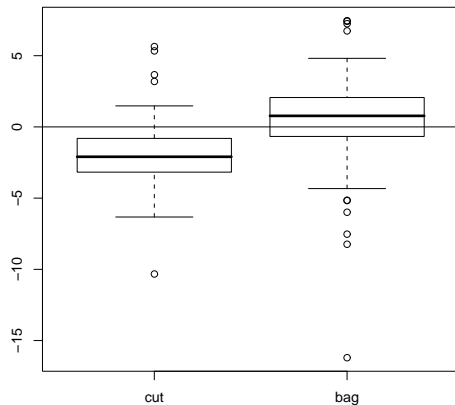


図 3: アルゴリズムの比較. 縦軸  $R$  の値. 左: データフィルタリングアルゴリズム, 右: TrBagg

- データをフィルタリングするしきい値  $\varepsilon = 0.1$

また, それぞれの条件で乱数の種を変えて 100 回ずつ実験を行い統計的に比較を行う.

**TrBagg との比較** まず, 上記のデフォルト条件で, TrBagg による方法とデータフィルタリングアルゴリズムの比較を行った. 理論解析で予測されるように, TrBagg の  $R$  値は 0 付近に平均値があり, データフィルタリングアルゴリズムの方は  $R$  が負となっている場合が多く解の改善が見られる (図 3).

**ターゲット分布の分散  $\sigma_T$  を変えたときの振る舞い** (16) 式によれば,  $\sigma_T$  をコントロールする  $c$  は小さいほどよいが,  $\sigma_T$  を大きくしていくとどうなるだろうか. 図 3 は  $\sigma_T$  を  $0.02 * 2^k$ ,  $k = 0, 1, 2, \dots, 5$  と変化させたときの結果である.  $\sigma_T$  が大きくなると改善の度合いは小さくなるが, 平均的に見れば悪くなることはない. また, 分散はだんだん小さくなっており  $\sigma_T$  が大きくなっても不安定にはならない様子が見られる.

**窓の幅  $\varepsilon$  を変えたときの振る舞い** 次に, データをフィルタリングするための窓の幅  $\varepsilon$  を変化させたときの様子を図 5 に示す<sup>3</sup>. 値は  $0.01 * 2^k$ ,  $k = 0, 1, \dots, 5$  に取った.  $\varepsilon$  は解析におけるコントロール変数であり, ほかのいろいろな量と関係しているため振る舞いは複雑である. 図からまずわかることは,  $\varepsilon$  を大きくして行くにつれて改善の度合いが増している様子である. これは相対

<sup>3</sup> $\varepsilon$  は解析ではコントロール変数だったが, ここでは窓の幅だけを変え, それ以外はデフォルト値に固定している.

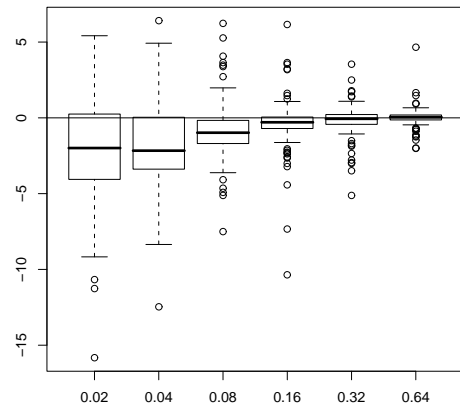


図 4: ターゲット分布の分散  $\sigma_T$  を変えたときの  $R$  の振る舞い

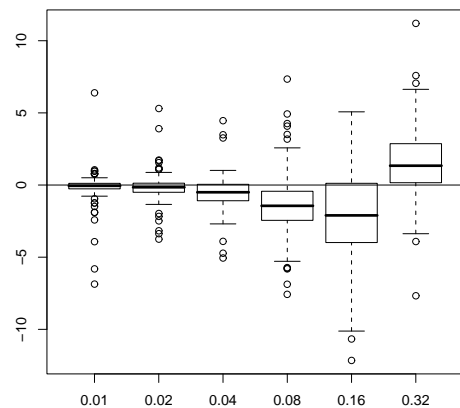


図 5: 窓の幅 ( $\varepsilon$ ) を変えたときの  $R$  の振る舞い

的に  $\sigma_T$  を小さくしていることに対応していると考えられる. ただし,  $\varepsilon$  をある点より大きくしすぎると今度は急な悪化が見られる. これは  $\varepsilon$  が微量量であるという仮定が崩れていることに相当するが, 定性的には  $p_N(x)$  の中心に近いサンプルが入ってくる影響と思われる.

**混合比  $\alpha = a\varepsilon$  を変えたときの振る舞い**  $S$  に含まれているサンプルのうち  $p_T(x)$  から生成されたものの数は多い方がよい. 図 6 に,  $|S| = 1000$  に保ったまま,  $p_T(x)$  由来のサンプルの個数を  $10 * 2^k$ ,  $k = 0, 1, 2, \dots, 5$  と変化させたとき (これはほぼ  $\alpha|S|$  に相当する) の結果を示す. 当然  $\alpha$  が大きいほど改善の度合いも大きい, 解析結果からも予測できるように,  $\alpha$  は  $\varepsilon$  オーダーの

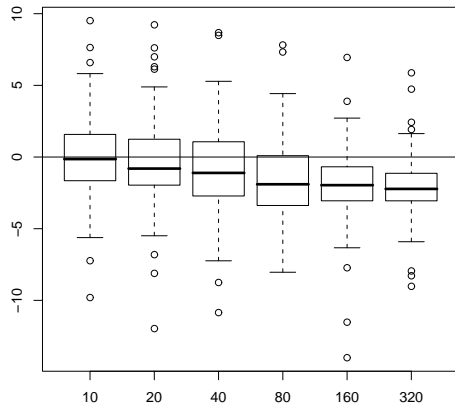


図 6: 混合比  $\alpha$  を変えたときの  $R$  の振る舞い (横軸は  $S$  中の  $p_T(x)$  由来のサンプル数で, ほぼ  $\alpha|S|$  に相当) 小さな場合でもそれほど悪い値にはならない.

## 5 一般の場合への拡張と問題点

さて, 前節で解説したモデルは非常に単純なものであったので, その結果がどの程度一般化できるかについては検討の必要がある.

まず,  $p_N(x)$  が単一の正規分布と仮定したことについては, 実際問題では雑多な種類のデータが混じっていることから問題があると感じられるかもしれないが, 雑多な種類のデータのうち最も  $p_T(x)$  に近いところに  $p_N(x)$  を集中させていると見れば, 一種の最悪評価とも考えられる. 解析結果から示唆されるのは,  $p_N(x)$  が  $p_T(x)$  の近くにピークをもつような場合には解の改善が難しいという点であり,  $|S|$  の中に  $p_T(x)$  と同程度の割合で,  $p_T(x)$  に非常に近い分布から出てきたサンプルがある場合にはそちらに解が引っ張られてしまうかもしれないということである. ただし,  $p_T(x)$  に近いという意味では逆にそれほど解が破綻するようなことはないとも考えられる.

次に, TrBagg が経験的に識別問題でうまく働くということであれば, 識別問題を単純化したモデルに対して解析を行う必要があるかもしれない. これについては今後の課題である.

## 6 おわりに

二つの品質の異なるデータ集合からの学習問題において, 単純な数理モデルを解析してその振る舞いを調べた. この問題を含め転移学習は, さまざまな応用をもつ実用的にも重要な問題であり, 比較的古くから研究され

てきたにも関わらず, 理論的な解析はそれほど進んでいないのが現状である. 今後も理論面での解析を進めると共に, 実データでの検証も並行して行っていきたい.

## 参考文献

- [1] 神島, 濱崎, 赤穂: “飼いならし 飼育・野生混在データからの学習”, 人工知能学会全国大会 (第 22 回) 論文集, 2D1-3, 2008.
- [2] 神島, 濱崎, 赤穂: “飼いならしを用いた協調タグ付けのタグ予測”, 2008 年度統計関連学会連合大会講演報告集, p.88, 2008
- [3] T. Kamishima, M. Hamasaki, S. Akaho: “Bag-taming. learning from wild and tame data,” in ECML/PKDD2008 Workshop: Wikis, Blogs, Bookmarking Tools. Mining the Web 2.0, 2008.
- [4] T. Kamishima, M. Hamasaki, S. Akaho: “Personalized tag prediction boosted by bagtaming.a case study of the hatena bookmark,” in the 3rd Int. Workshop on Data-Mining and Statistical Science, 2008.
- [5] T. Kamishima, M. Hamasaki, S. Akaho: “TrBagg: A Simple Transfer Learning Method and Its Application to Personalization in Collaborative Tagging,” in Proc. IEEE Int. Conf. on Data Mining (ICDM2009), to appear.
- [6] S. J. Pan, Q. Yang: “A Survey on Transfer Learning,” Technical Report, Dept. of Computer Science and Engineering, Hong Kong Univ. of Science and Technology, HKUST-CS08-08, 2008.