3B1-R-2-5

Density power divergenceを用いたロバスト能動学習

Robust Active Learning via Density power divergence

十河 泰弘[*1] 植野 剛[*2] 河原 吉伸[*1*2] 鷲尾 隆[*1*2]
Yasuhiro Sogawa Tsuyoshi Ueno Tsuyoshi Ueno Tsuyoshi Ueno

[*1]大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

[*2]科学技術振興機構
Japan Science and Technology Agency

The accuracy of active learning is critically influenced by the existence of outliers in input samples. In this paper, we propose a novel pool-based active learning framework through robust measures based on density power divergence. It is known that density power divergence, such as β-divergence and γ-divergence, can be accurately estimated even under the existence of outliers within data. Hence, we develop query selecting measures for pool-based active learning using these divergences. Also, we propose an evaluation scheme for these measures based on those asymptotic statistical analyses, which enables us to perform active learning by evaluating an estimation error directly. Experiments with real-world image datasets show that our active learning scheme performs better than several state-of-the-art methods.

1. はじめに

近年では計算機やストレージの発達に伴い，様々な分野において莫大なデータを容易に蓄積することが可能となった．

しかし，それらのデータの中で大多数を占めているのはラベルなし事例であり，ラベル有り事例は少数に限られている．これは専門家や実験によるラベル付けが莫大な費用・時間を要することに起因している．そのため，これらのデータに対してラベル有り事例を必要とする手法を適用した際に，ラベル有り事例の不足ゆえに十分なパフォーマンスを得られないケースが多く存在している．これらの背景の下で，近年ではこのような大多数ラベル無し事例と少数ラベル有り事例を同時に活用する能動学習と呼ばれる手法が盛んに研究されている [Campbell 00, Tong 02, Kanamori 03]．能動学習は単にラベル有り事例を与えられて学習を行う受動的な学習手法と異なり，計算機側が，ラベル獲得によって最も学習精度向上が期待される事例を選択（クエリ）し，ユーザーがそのクエリ事例に対してラベル付けを行うことで，できる限り少数のラベル有り事例を用いて高いパフォーマンスを得ることを目的とした手法である．この能動学習は音声認識や分類問題などの様々な応用に適用され，成功を収めている [Hakkani-Tur 02, McCallum 98]．

能動学習において最も重要な問題は，クエリ事例の選択基準であり，ここ数年において様々なクエリ基準を用いた手法が提案されてきた [Settles 08]．これらの既存手法に共通して言えることはいずれも，オラクル (専門家や追加実験) が与えるラベル付けは必ず正しい，という強い仮定を置いていることである．しかしながら現実問題を扱う際には，むしろそのようなケースは稀である．専門家によるラベル付けはその人の疲労等の状態によってはラベル付けを誤る場合があり，また，実験によるラベル付けも実験環境によってラベル付けを誤る可能性がある．このように誤ったラベル付けを行うオラクルは Noisy oracle と呼ばれ，それによるラベル付けは最終的な能動学習のモデル推定結果を悪化させてしまう．そこで本研究では，ラベル付けが常に正しいという仮定を緩和し，オラクルがラベル付けを誤る場合を想定した能動学習手法の提案を行う．

連絡先: 十河泰弘，大阪府茨木市美穂ケ丘８－１ 産業科学研究所 知能推論研究分野，sogawa@ar.sanken.osaka-u.ac.jp

先に述べたようにクエリ指標にも様々なものが存在しているが，本研究では Expected error reduction approach と呼ばれる，モデルの汎化誤差に基づく指標を用いる．汎化誤差の指標にはカルバック・ライブラー・ダイバージェンス (KL-ダイバージェンス) が一般的に用いられるが，本論文では KL-ダイバージェンスと同様のクラスに属し，より外れ値にロバストに機能することが知られる，β-ダイバージェンスと，同様にロバストな γ-ダイバージェンスを用いる．これらのダイバージェンスは Density power divergence として知られている．従来の期待誤差を最小化する方策では，クエリ選択を行う際に繰り返しモデルの学習を行う必要があったが，我々のロバストダイバージェンスの漸近解析に基づく能動学習手法では，それが不要であり，計算コストが小さく済む．本論文では，我々のクエリ戦略を 2 値クラス分類問題を解くうえでよく知られるロジスティックモデルに導入し，実データを用いた評価実験によって提案手法の学習精度の検証を行った．

2. 能動学習とダイバージェンス

2.1 プールベース能動学習

本研究ではプールベース能動学習と呼ばれる，データの真のテスト分布はわかっていないものの，その分布から多くの事例が得られているという前提の下で議論を行う．このプールベース能動学習のアルゴリズムの全体像を Algorithm 1 に示す．プールベース能動学習では，まず最初に少数ラベル有り事例集合 $\mathcal{L}$ からパラメータ $\boldsymbol{\theta}$ の推定を行い，モデル $p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)$ を得る (Algorithm 1, LearnModel)．次にその確率モデルを基に最も’興味深い’事例をラベル無し事例集合 $\mathcal{U}$ から選択し，クエリする (Algorithm 1, SelectQuery)．’興味深い’事例としては一般にモデルの学習に対して最も精度向上が期待される事例が選択される．そして，そのクエリ事例に対してオラクル（専門家や追加実験）がラベル付けを行い，それをラベル有り事例集合に加えて再度パラメータの推定を行う．これを繰り返して，最終的に推定モデルを得る．冒頭でも述べたように，能動学習においてクエリ指標の決定は重要な問題である．クエリ指標の１つとしては，最良のモデル $p_{\bar{\boldsymbol{\theta}}}(\mathbf{x}, y)$ と推定モデル $p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)$ との間の誤差の最小化に基づく方法が挙げられる．ここで述べた最良のモデルとは，ラベル有り事例が無限に与え

Algorithm 1 プールベース能動学習

- **Input**
  - $\mathcal{U}$: ラベル無し事例集合
  - $\mathcal{L}$: ラベル有り事例集合
  - K:1 度のクエリで選択される事例数
  - T:クエリ回数
- **Main**
  - $\boldsymbol{\theta}^{(0)}$ = LearnModel($\mathcal{L}$)
  - for i=1,$\cdots$, T
    - $\mathcal{S}$ = SelectQuery($\mathcal{U}$,K,$\boldsymbol{\theta}^{(i-1)}$)
    - $\mathcal{L} = \mathcal{L} \cup \mathcal{S}$
    - $\mathcal{U} = \mathcal{U} \backslash \mathcal{S}$
    - $\boldsymbol{\theta}^{(i)}$ = LearnModel($\mathcal{L}$)
  - end
- **Output**
  - $\boldsymbol{\theta}^{(T)}$: 学習されたパラメータ

られた際に真の分布 $q(\mathbf{x}, y)$ との間のダイバージェンスを最小化するパラメータ $\hat{\boldsymbol{\theta}}$ に基づくモデル $p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, y)$ を指す．y はラベルであり，この論文においては 2 値 $y \in \{0, 1\}$ である．この誤差最小化に基づくクエリ選択戦略は Expected error reduction approach[Roy 01] として知られている．しかし，この KL-ダイバージェンスを用いた方法はラベルノイズを考慮していないため，本論文では次項で説明するロバストなダイバージェンスを用いることを提案する．

2.2 Density power divergence

これまで述べたように，従来の Expected error reduction approach に対して，本研究では Density power divergence と呼ばれるクラスに属するロバストダイバージェンスを導入する．ロバストダイバージェンスのうち，特に本研究では β-ダイバージェンスと γ-ダイバージェンスに着目し，本節ではその詳細を簡単に説明する．

2.2.1 β-ダイバージェンス

Density power divergence は 2 つの分布の差異を測る統計的尺度のクラスであり，外れ値を含むデータからロバスト推定を行うことを目的として提案された．その尺度の 1 つである β-ダイバージェンスは [Basu 98] で提案され，次のように定義されている．

$$D_\beta(q\|p_{\boldsymbol{\theta}}) = \frac{1}{(1+\beta)}\left\{\frac{1}{\beta}\iint q(\mathbf{x},y)^{1+\beta}\mathrm{d}\mathbf{x}\mathrm{d}y - \iint q(\mathbf{x},y)p_{\boldsymbol{\theta}}(\mathbf{x},y)^{\beta}\mathrm{d}\mathbf{x}\mathrm{d}y + \iint p_{\boldsymbol{\theta}}(\mathbf{x},y)^{1+\beta}\mathrm{d}\mathbf{x}\mathrm{d}y\right\}.$$

ここでの β は正の定数である．$\beta \to 0$ のとき，β-ダイバージェンスは KL-ダイバージェンスに収束する．そのため，この β-ダイバージェンスは KL-ダイバージェンスの一般化したものと捉えることができる．

他のダイバージェンスと同様に，この β-ダイバージェンスを最小化することによってパラメータの推定を行うことが可能である．ここで，事例集合 $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ と，それら事例のラベル $\{y_1, \cdots, y_n\}$ が真の分布 $q(\mathbf{x}, y)$ から得られたとすると，このときの β-ダイバージェンスに基づくモデルのパラメータ

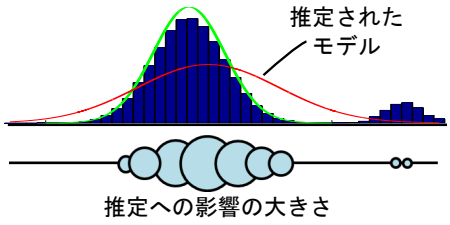

図 1: 重み付き推定の例

は以下の推定方程式の解として与えられる．

$$\sum_{i=1}^{n} p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i, y_i)^{\beta}\partial_{\boldsymbol{\theta}} \ln p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i, y_i) - \iint p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)^{\beta+1}\partial_{\boldsymbol{\theta}} \ln p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)\mathrm{d}\mathbf{x}\mathrm{d}y = \mathbf{0}. \quad (1)$$

この推定方程式は β-ダイバージェンスを微分し，$q(\mathbf{x}, y)$ に関する期待値計算を標本平均によって置き換えることで得られる．ここで，$\partial_{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ に関する偏微分を表す．Density power divergence に共通な性質は，いずれも Eq. (1) のように自身の分布で尤度関数に重み付けを行っていることであり，この重み付けにより外れ値を無視した推定が可能となっている．図 1 は実際に重み付き推定方程式がどのように外れ値に対してロバストな推定を行っているかの一例である．図中の緑線は真の分布，赤線は外れ値の影響化での分布を表している．この図のように，一般的に外れ値はモデル $p_{\boldsymbol{\theta}}$ において非常に小さな確率で得られるため，重み付き尤度方程式において外れ値が推定に与える影響も非常に小さなものとなる．このように Density power divergence を用いた場合，外れ値に対してロバストにパラメータ推定を行うことが可能となる．

2.2.2 γ-ダイバージェンス

β-ダイバージェンスの改良型の 1 つである，γ-ダイバージェンスは以下のように定義される [Fujisawa 08].

$$D_\gamma(q\|p_{\boldsymbol{\theta}}) = \frac{1}{\gamma+1}\left\{\frac{1}{\gamma}\ln\iint q(\mathbf{x},y)^{1+\gamma}\mathrm{d}\mathbf{x}\mathrm{d}y - \ln\iint q(\mathbf{x},y)p_{\boldsymbol{\theta}}(\mathbf{x},y)^{\gamma}\mathrm{d}\mathbf{x}\mathrm{d}y + \ln\iint p_{\boldsymbol{\theta}}(\mathbf{x},y)^{1+\gamma}\mathrm{d}\mathbf{x}\mathrm{d}y\right\}.$$

同様にして，γ-ダイバージェンスに基づく推定方程式も得られる．

$$\sum_{i=1}^{n}\left(\frac{p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i, y_i)^{\gamma}}{\sum_{i=1}^{n} p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i, y_i)^{\gamma}}\right)\partial_{\boldsymbol{\theta}} \ln p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i, y_i) - \iint\left(\frac{p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)^{\gamma+1}}{\iint p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)^{\gamma+1}\mathrm{d}\mathbf{x}\mathrm{d}y}\right)\partial_{\boldsymbol{\theta}} \ln p_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}, y)\mathrm{d}\mathbf{x}\mathrm{d}y = \mathbf{0}. \quad (2)$$

この式は，Eq. (1) を正規化したものと捉えることができる．この推定方程式は β-ダイバージェンスのものとよく似ているが，β-ダイバージェンスと比べ，外れ値に対してよりロバストなパラメータ推定が可能であることが知られている．それゆえ，γ-ダイバージェンスは我々の提案するノイズラベル有り能動学習において，非常に効果的に機能すると期待される．

3. リスク関数の漸近解析に基づくクエリ指標

本節では我々のロバスト能動学習におけるクエリ指標の基礎となる，リスク関数（ダイバージェンス）に対する漸近解析の

結果を示す．既存の能動学習手法では，モデルが真の分布を含んでおり，正しく表現可能であるという仮定の下で研究が行われている．しかし，本研究ではその仮定を置かずに漸近解析を行い，その結果に基づく β/γ-ダイバージェンスによる期待誤差を用いたクエリ指標を提案する．

3.1 期待誤差の分解

ここでは特定のダイバージェンスやその推定量に限定せず，一般的な M 推定量 [Huber 09] に対する漸近解析を行う．最初に，$f(\boldsymbol{\theta}) = \mathbb{E}_q[\ell(\mathbf{x}, y; \boldsymbol{\theta})]$ という関数を定義する．ここで，$\ell(\mathbf{x}, y; \boldsymbol{\theta})$ は任意の関数であり，パラメータ $\boldsymbol{\theta}$ の良さの尺度である．また，$\mathbb{E}_q[\cdot]$ は $q(\mathbf{x}, y)$ に関する期待値を表す．ラベル有り事例集合 $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ が与えられたとき，この $f(\boldsymbol{\theta})$ を最小化する最適なパラメータ $\hat{\boldsymbol{\theta}}_n$ は

$$\sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}_n) = \mathbf{0}, \tag{3}$$

を解くことによって得られる．ここで，$\psi(\mathbf{x}, y; \boldsymbol{\theta})$ は $\ell(\mathbf{x}, y; \boldsymbol{\theta})$ を $\boldsymbol{\theta}$ に関して偏微分したもので，$\boldsymbol{\psi}(\mathbf{x}, y; \boldsymbol{\theta}) := \partial_{\boldsymbol{\theta}} \ell(\mathbf{x}, y; \boldsymbol{\theta})$ である．Eq. (3) を解くことで得られる推定量がいわゆる M 推定量である．Eq. (3) で与えられる推定量のクラスには最尤推定量や本研究で扱う β/γ-ダイバージェンスに基づく推定量も含まれる．そのため，本論文では M 推定量の解析を通じて，種々の推定量を一括して説明する．

本論文では，モデルの期待誤差を $f(\hat{\boldsymbol{\theta}}_n) = \mathbb{E}_q[\ell(\mathbf{x}, y; \hat{\boldsymbol{\theta}}_n)]$ と定義する．$n \to \infty$ のとき，推定量 $\hat{\boldsymbol{\theta}}_n$ が $\bar{\boldsymbol{\theta}}$ に確率収束すると仮定すると，$f(\hat{\boldsymbol{\theta}}_n)$ は以下のように分解できる．

$$f(\hat{\boldsymbol{\theta}}_n) = \underbrace{f(\hat{\boldsymbol{\theta}}_n) - f(\bar{\boldsymbol{\theta}})}_{\text{推定誤差}} + \underbrace{f(\bar{\boldsymbol{\theta}})}_{\text{近似誤差}}. \tag{4}$$

ここでの推定誤差は事例数が有限であることに起因するものである．一方で，近似誤差は推定量の収束点における誤差であり，モデルの表現能力に起因する．したがって，モデルが与えられた場合にこれは定数となるため，本研究では，サンプルの影響を受ける推定誤差にのみ着目し，これを最も減少させるようクエリ選択を行う．

3.2 推定誤差の漸近解析とクエリ指標

本項では，Eq. (4) で定義した推定誤差について述べる．スペースの都合上，漸近解析に基づく推定誤差の計算結果のみを記述する．いくつかの一般的な仮定の下で推定誤差は下記として得られる．

$$\mathbb{E}_q[f(\hat{\boldsymbol{\theta}}_n) - f(\bar{\boldsymbol{\theta}})] = \frac{1}{2n} \mathrm{tr}\left\{ (\mathbf{A}_q^{-1})^\top \mathbf{M}_q \right\} + o\left(\frac{1}{n}\right). \tag{5}$$

ここで，$\mathbf{A}_q, \mathbf{M}_q$ はそれぞれ，

$$\mathbf{A}_q := \mathbf{A}_q(\bar{\boldsymbol{\theta}}) := \mathbb{E}_q\left[\partial_{\boldsymbol{\theta}} \boldsymbol{\psi}(\mathbf{x}, y; \bar{\boldsymbol{\theta}})\right] \tag{6}$$

$$\mathbf{M}_q := \mathbf{M}_q(\bar{\boldsymbol{\theta}}) := \mathbb{E}_q\left[\boldsymbol{\psi}(\mathbf{x}, y; \bar{\boldsymbol{\theta}}) \boldsymbol{\psi}(\mathbf{x}, y; \bar{\boldsymbol{\theta}})^\top\right], \tag{7}$$

である．漸近解析については [Vaart 00] が詳しい．$f(\boldsymbol{\theta})$ が KL-ダイバージェンスである場合には，Eq. (5) は [Kanamori 03] にある結果と一致する．

ここまでの計算結果を元に下記の形で表されるロジスティックモデル上での能動学習クエリ指標へ導入する．

$$p(y|\mathbf{x}; \boldsymbol{\theta}) = \frac{\exp(y \boldsymbol{\theta}^\top \mathbf{x})}{1 + \exp(\boldsymbol{\theta}^\top \mathbf{x})}. \tag{8}$$

本来ならば Eq. (6) ならびに (7) を計算する必要があるが，真の分布 $q(\mathbf{x}, y)$ ならびに収束パラメータ $\bar{\boldsymbol{\theta}}$ を直接知ることはできない．そこで，これらの式を計算するために，実用上では，現段階においてラベルあり事例からの推定したモデル $p_{\hat{\boldsymbol{\theta}}_n}(y|\mathbf{x})$ による近似と，標本平均を用いて次のように計算を行う．

$$\widehat{\mathbf{A}}_{p_{\hat{\boldsymbol{\theta}}_n}}(\mathcal{S}) = \sum_{\mathbf{x}_i \in \mathcal{S}} \sum_{y \in \{0,1\}} p_{\hat{\boldsymbol{\theta}}_n}(y|\mathbf{x}_i) \partial_{\boldsymbol{\theta}} \boldsymbol{\psi}(y|\mathbf{x}_i; \hat{\boldsymbol{\theta}}_n)$$

$$\widehat{\mathbf{M}}_{p_{\hat{\boldsymbol{\theta}}_n}}(\mathcal{S}) = \sum_{\mathbf{x}_i \in \mathcal{S}} \sum_{y \in \{0,1\}} p_{\hat{\boldsymbol{\theta}}_n}(y|\mathbf{x}_i) \boldsymbol{\psi}(y|\mathbf{x}_i; \hat{\boldsymbol{\theta}}_n) \boldsymbol{\psi}(y|\mathbf{x}_i; \hat{\boldsymbol{\theta}}_n)^\top.$$

ここで，$\mathcal{S}$ はクエリとして選択されたラベル無し事例集合であり，$\psi(\cdot)$ はロジスティックモデルにおける β-ダイバージェンス，もしくは γ-ダイバージェンスの推定関数と一致する．

前節において説明した Algorithm 1 の SelectQuery のステップにおいてクエリ集合 $\mathcal{S}$ は

$$\mathcal{S}^* = \underset{\mathcal{S} \subseteq \mathcal{U} \wedge |\mathcal{S}| = K}{\arg\min} \frac{1}{2K} \mathrm{tr}\left\{ \left(\hat{\mathbf{A}}_{p_{\hat{\boldsymbol{\theta}}_n}}(\mathcal{S})^{-1}\right)^\top \hat{\mathbf{M}}_{p_{\hat{\boldsymbol{\theta}}_n}}(\mathcal{S}) \right\},$$

を満たすように決定される．ここで，K は一度にクエリされる事例数である．$K = 1$ のときは，本能動学習手法はいわゆるシングルモード能動学習であり，十分なラベル有り事例を得るために，繰り返しモデルの学習とクエリ選択を行う必要がある．しかし，モデルの学習に多くの計算時間を費やすため，実用上ではクエリ時に複数の事例を一括で選択するバッチモード能動学習が好まれる．そのため，本研究ではバッチモード能動学習を採用する．クエリ集合 $\mathcal{S}$ の最適化は事例集合における組み合わせ最適化問題であり，容易に解くことができないが，能動学習の枠組みにおいては多くの場合に貪欲法を用いるため，本研究でも同様に貪欲法を用いて集合 $\mathcal{S}$ の最適化を行う．

4. 評価実験

本節では前項までで提案したクエリ指標を基にした 2 つの提案手法と 3 つの既存手法との比較実験の結果を示す．

1. β-AL: β-ダイバージェンスに基づくクエリ指標を用いた提案手法．
2. γ-AL: γ-ダイバージェンスに基づくクエリ指標を用いた提案手法．
3. BMAL: フィッシャー情報量に基づいたクエリ指標を用いる既存手法 [Hoi 06]．
4. β-RAND: ランダムにクエリ選択を行い，β-ダイバージェンスに基づくパラメータ推定を行う手法．
5. γ-RAND: ランダムにクエリ選択を行い，γ-ダイバージェンスに基づくパラメータ推定を行う手法．

BMAL はパラメータの漸近分散に基づいた能動学習手法であり，漸近解析に基づく本研究と近い．また，本研究と同様にバッチ能動学習を採用している．これらの理由から，能動学習手法の比較対象として BMAL を選択した．評価実験には 0 から 9 までの人の手書き数字画像を収集した MNIST の手書き数字データを用いた [LeCun 98]．その中でも特に形が似ており判別が難しいと考えられる，1 と 7 の画像データの分類による評価実験を行った．本データは本来 784 次元 65000 事例からなるデータであるが，前処理として一般的な主成分分析を用

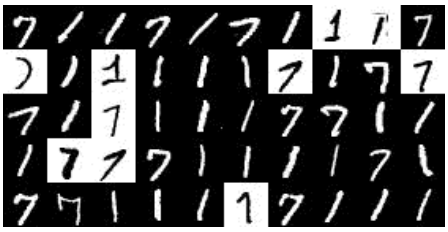

図 2: クエリされた事例の一例

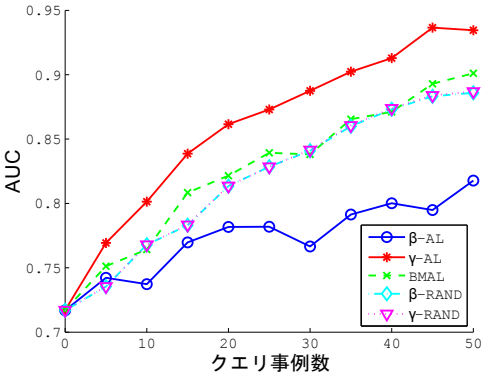

図 3: MNIST 手書き数字データの分類精度の比較

いて特徴次元を 20 次元まで削減し，65000 事例のうち，ランダムに 100 事例を選択して，実験を行った．実験に当たっては、最初に 80 の訓練事例と 20 のテスト事例にランダムに分割した．そして，初期ラベル有り事例数を 6，一度にクエリする事例数を 5，クエリ回数を 10 回とし，パラメータ $\beta, \gamma = 0.01$ とした．また，クエリされた事例に対してユーザが 1 か 7 かのラベル付けを行った際のラベル付け誤り率は約 15 %であった．図 2 は実際に用いたデータの一部で，白背景に黒文字がラベル付けを誤った事例である．この図 2 より，似た文字のラベル付けを誤っていることが伺える．この手書き数字データを提案手法と比較手法にそれぞれ適用し，その推定精度を AUC を用いて評価した．そして，この試行を 50 回繰り返し，AUC の平均をとったところ図 3 の結果を得た．図 3 はクエリ事例追加ごとの学習精度の変化を表しており，いずれの手法に関してもクエリの増加に伴って精度が上昇していることがわかる．また，この図から，MNIST 手書き数字データを用いた実験において γ-AL は従来手法よりも精度よく推定できていることがわかる．一方，β-AL は精度が低いが，これは正規化を用いている γ-ダイバージェンスに比べ，β-ダイバージェンスがパラメータ β の影響を受けやすいためであると考えられる．

5. まとめ

本研究では，一般的なリスク関数の漸近解析を行い，それに対して Density power divergence を導入することで，Noisy oracle に対してロバストな能動学習手法の提案を行った．さらに本研究で提案したロバスト能動学習手法の枠組みをロジスティックモデルに導入し，実データを用いた評価実験によって提案手法の実用性を示した．本研究によるクエリ戦略は，ロジスティックモデルと同様に他のモデルへも導入が可能であると考えられるため，その際の挙動の確認が今後の課題として挙げられる．

参考文献

[Basu 98] Basu, A., Harris, I., Hjort, N., and Jones, M.: Robust and efficient estimation by minimising a density power divergence, *Biometrika*, Vol. 85, No. 3, pp. 549–559 (1998)

[Campbell 00] Campbell, C., Cristianini, N., and Smola, A.: Query learning with large margin classifiers, in *Proceedings of the 17th International Conference on Machine Learning*, pp. 111–118 (2000)

[Fujisawa 08] Fujisawa, H. and Eguchi, S.: Robust parameter estimation with a small bias against heavy contamination, *Journal of Multivariate Analysis*, Vol. 99, No. 9, pp. 2053–2081 (2008)

[Hakkani-Tur 02] Hakkani-Tur, D., Riccardi, G., and Gorin, A.: Active learning for automatic speech recognition, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 3904–39078 (2002)

[Hoi 06] Hoi, S., Jin, R., Zhu, J., and Lyu, M.: Batch mode active learning and its application to medical image classification, in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 417–424 (2006)

[Huber 09] Huber, P. J. and Ronchetti, E. M.: *Robust statistics*, John Wiley and Sons (2009)

[Kanamori 03] Kanamori, T. and Shimodaira, H.: Active learning algorithm using the maximum weighted log-likelihood estimator, *Journal of Statistical Planning and Inference*, Vol. 116, No. 1, pp. 149–162 (2003)

[LeCun 98] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (1998)

[McCallum 98] McCallum, A. and Nigam, K.: Employing EM in pool-based active learning for text classification, in *Proceedings of the 15th International Conference on Machine Learning*, pp. 350–358 (1998)

[Roy 01] Roy, N. and McCallum, A.: Toward optimal active learning through sampling estimation of error reduction, in *Proceedings of the 18th International Conference on Machine Learning*, pp. 441–448 (2001)

[Settles 08] Settles, B., Craven, M., and Ray, S.: Multiple-instance active learning, in *Advances in Neural Information Processing Systems 20* (2008)

[Tong 02] Tong, S. and Koller, D.: Support vector machine active learning with applications to text classification, *Journal of Machine Learning Research*, Vol. 2, pp. 45–66 (2002)

[Vaart 00] Vaart, Van der A.: *Asymptotic statistics*, Cambridge Univ Pr (2000)