Generalized Bradley-Terry モデルを用いた 二値判別器の組み合わせによる多値判別法

1X08C033-1　荻原大陸
指導教員　後藤正幸

1　研究背景と目的

近年，コンピュータネットワークの発達に伴い，電子文書が大量に扱われるようになった．これらの電子文書は，情報の膨大さから人手による分類が難しく，自動分類技術が望まれている．そのような背景から，Support Vector Machine (SVM) や Relevance Vector Machine (RVM) のように性能の良い二値判別器による自動分類手法が提案されている．多値判別問題に対しては，直接多値判別器を構成するよりも，二値判別器を組み合わせて多値判別を行う方が効率的であるため，二値判別器の組み合わせによる多値判別器の構成法が多く提案されている [1].

本研究では複数の二値判別器の組み合わせによる多値判別において，様々な拡張性が望める Generalized Bradley-Terry(GBT) モデル [2] を用いた判別器の構成法に焦点を当てる．GBT モデルは Bradley-Terry(BT) モデルを基に，任意の判別器の組み合わせを可能にしたものである [3]．一般的な多値判別問題では，入力データと分類カテゴリの関係によって，判別器が分類し易いカテゴリや分類しにくいカテゴリが混在するため，二値判別器の精度にばらつきが生じる．しかし，GBT モデルによる判別器の構成法では，それぞれの二値判別器の精度の差異を全く考慮せずに多値判別に用いているため，精度の悪い判別器が精度の良い判別器と同等の影響を与え，判別性能の低下につながっている可能性がある．この問題を解決するため，各二値判別器の精度を導出して重み付けを行うことで，精度のばらつきを考慮した多値判別手法を提案する．提案手法の有効性を示すため，新聞記事を用いて分類実験を行い，分類精度が向上することを検証する．

2　準備

2.1　多値判別問題

K をカテゴリ数，カテゴリの集合を $\mathcal{C}=\{c_1,\ldots,c_K\}$ とする．判別問題とはカテゴリが既知の学習データを使って学習を行い，カテゴリが未知の新たな判別対象データ (入力ベクトル) $\boldsymbol{x}$ に対応するカテゴリ $c\in\mathcal{C}$ を推定することである．多値判別問題とは $K>2$ の場合の判別問題を指し，一方，二値判別問題は $K=2$ の場合の判別問題を指す.

2.2　二値判別器

本研究では，二値判別器として精度が良いとされる SVM の特性を多く保持し，各カテゴリに属する確率を推定して軟判定を行う RVM を用いる．判別器の個数を R とし，二値判別器 r $(r=1,\ldots,R)$ は，入力ベクトル $\boldsymbol{x}$ が $\mathcal{C}_r^+$ のカテゴリ集合か $\mathcal{C}_r^-$ のカテゴリ集合のどちらに属するかを判別する．ここで，$\mathcal{C}_r^+,\mathcal{C}_r^-\subset\mathcal{C}$, $\mathcal{C}_r^+,\mathcal{C}_r^-\neq\phi$, $\mathcal{C}_r^+\cap\mathcal{C}_r^-=\phi$, $\mathcal{C}_r=\mathcal{C}_r^+\cup\mathcal{C}_r^-$, とする．カテゴリの分割が $|\mathcal{C}_r^+|=|\mathcal{C}_r^-|=1$ であるならば 1-vs-1 判別器と呼ばれ，$|\mathcal{C}_r^+|=1$, $|\mathcal{C}_r^-|=K-1$ ならば 1-vs-the rest 判別器と呼ばれる．二値判別器 r は確率 $q_r(\boldsymbol{x})$ を返すため，判別器が示す確信度として用いることが出来る．なお，$q_r(\boldsymbol{x})$, $1-q_r(\boldsymbol{x})$ は，それぞれ $P(c\in\mathcal{C}_r^+|c\in\mathcal{C}_r,\boldsymbol{x})$, $P(c\in\mathcal{C}_r^-|c\in\mathcal{C}_r,\boldsymbol{x})$ の推定値と見なすことが出来る．

3　従来手法

3.1　BT モデル

BT モデルは多数のプレーヤーが 1 対 1 の試合を多く行ったとき，各プレーヤーの強さを定量化するモデルである．プレーヤーが K 人存在し，プレーヤー k $(k=1,\ldots,K)$ は強さと呼ばれる非負のパラメータ p_k を持つと仮定する．また，プレーヤー k がプレーヤー ℓ $(\ell=1,\ldots,K,\ell\neq k)$ に勝つ確率を $\alpha_{k\ell}=p_k/(p_k+p_\ell)$, プレーヤー ℓ がプレーヤー k に勝つ確率を $\alpha_{\ell k}=p_\ell/(p_k+p_\ell)$ と仮定する．プレーヤー k と ℓ の試合数を $n_{k\ell}$ とし，その時の k の勝率を $r_{k\ell}=$ (k が ℓに勝った回数)$/n_{k\ell}$ で定義する．ここで引き分けはなし，すなわち $r_{k\ell}+r_{\ell k}=1$ が成り立つとする．この時，対数尤度関数 $F(\boldsymbol{p})$ を次式で定義する．

$$F(\boldsymbol{p})=\sum_{k=1}^{K}\sum_{\ell=k+1}^{K}n_{k\ell}\Big(r_{k\ell}\ln\alpha_{k\ell}+(1-r_{k\ell})\ln\alpha_{\ell k}\Big). \quad (1)$$

ここで，$\boldsymbol{p}=(p_1,\ldots,p_K)$ であり，$\boldsymbol{p}$ に関して $F(\boldsymbol{p})$ の最大化を行い，最尤推定量 $\hat{\boldsymbol{p}}=(\hat{p}_1,\ldots,\hat{p}_K)$ を得る．ただし，$\sum_{k=1}^{K}\hat{p}_k=1$, $\hat{p}_k>0$ である．得られた $\hat{\boldsymbol{p}}$ の各成分がプレーヤーの強さを表し，プレーヤーの順位付けに用いることが出来る．

3.2　BT モデルによる多値判別器の構成

二値判別器を組み合わせる手法に BT モデルを適用することを考える．BT モデルにおける (1) 式の $r_{k\ell}$ を 1-vs-1 判別器の出力 $q_r(\boldsymbol{x})$ に変更し，$n_{k\ell}=1$ とおく．k_r を $\mathcal{C}_r^+$ のカテゴリ番号，ℓ_r を $\mathcal{C}_r^-$ のカテゴリ番号とし，$\alpha_{k_r\ell_r}=p_{k_r}/(p_{k_r}+p_{\ell_r})$, $\alpha_{\ell_r k_r}=p_{\ell_r}/(p_{k_r}+p_{\ell_r})$ とした場合，BT モデルに二値判別器を適用させた対数尤度関数 $F_{\mathrm{BT}}(\boldsymbol{p},\boldsymbol{x})$ は次式のように定義される．

$$F_{\mathrm{BT}}(\boldsymbol{p},\boldsymbol{x})=\sum_{r=1}^{R}\Big(q_r(\boldsymbol{x})\ln\alpha_{k_r\ell_r}+(1-q_r(\boldsymbol{x}))\ln\alpha_{\ell_r k_r}\Big). \quad (2)$$

BT モデルを用いた多値判別器では従来の BT モデルと同様に，$\boldsymbol{p}$ に関して $F_{\mathrm{BT}}(\boldsymbol{p})$ の最大化を行い，最尤推定量 $\hat{\boldsymbol{p}}=(\hat{p}_1,\ldots,\hat{p}_K)$, $\sum_{k=1}^{K}\hat{p}_k=1$, $\hat{p}_k>0$ を得る．得られた $\hat{p}_k$ を $p(c_k|\boldsymbol{x})$ の推定値と見なす．$\hat{p}_k$ がカテゴリ c_k に所属する確率を表すため，これを用いて入力 $\boldsymbol{x}$ が所属するカテゴリを $\hat{c}=\mathrm{argmax}_{c_k\in\mathcal{C}}\hat{p}_k$ と推定することが出来る．

3.3　GBT モデルによる多値判別器の構成

BT モデルでは多値判別器を構成するために 1-vs-1 判別を用いているので，1-vs-the rest 判別に比べ，1 つの二値判別器の学習に用いるデータ数が少なくなり，精度の良い判別を難しくしている可能性がある．一方，GBT モデルでは，1-vs-1 判別器に加え，1-vs-the rest 判別器など任意の判別器

の組み合わせが可能である．GBT モデルに対する対数尤度関数 $F_{\mathrm{GBT}}(\boldsymbol{p},\boldsymbol{x})$ は次式のように定義される．

$$F_{\mathrm{GBT}}(\boldsymbol{p},\boldsymbol{x}) = \sum_{r=1}^{R}\Bigg(q_r(\boldsymbol{x})\ln\frac{\sum_{c_k\in\mathcal{C}_r^+} p_k}{\sum_{c_\ell\in\mathcal{C}_r} p_\ell} + (1-q_r(\boldsymbol{x}))\ln\frac{\sum_{c_k\in\mathcal{C}_r^-} p_k}{\sum_{c_\ell\in\mathcal{C}_r} p_\ell}\Bigg). \quad (3)$$

GBT モデルを用いた多値判別器でも同様に，$\boldsymbol{p}$ に関して $F_{\mathrm{GBT}}(\boldsymbol{p},\boldsymbol{x})$ の最大化を行い，最尤推定量 $\hat{\boldsymbol{p}}=(\hat{p}_1,\ldots,\hat{p}_K)$, $\sum_{k=1}^{K}\hat{p}_k=1$, $\hat{p}_k>0$ を得る．得られた $\hat{p}_k$ を $p(c_k|\boldsymbol{x})$ の推定値と見なす．

4　提案手法

本研究では最も基本的な判別器構成である 1-vs-the rest 判別器の組み合わせを対象とする．従来の GBT モデルによる多値判別では，すべての二値判別器の精度を同等と見なしているため，有効な判別が行えない場合がある．そこで提案手法では，判別器の精度を算出し，それを基に判別器 r の重みを決定する．得られた重みを GBT モデルに適用させることで判別器の精度のばらつきを考慮した GBT モデル (MGBT モデル) を構成する．

4.1　判別器ごとの重みの導出法

判別器の精度のばらつきは，判別器毎の信頼性の差異とも考えることができる．このことから，判別器の信頼性を比較する際には，判別器から得られる判別の確信度 $q_r(\boldsymbol{x})$ を指標とする．学習させた二値判別器の $\mathcal{C}_r^+$ に含まれる正例の学習データに対する出力を用いて判別器精度を導出する．学習データ集合を $\mathcal{X}'$, 学習データ $\boldsymbol{x}'(\boldsymbol{x}'\subset\mathcal{X}')$ のカテゴリを $c(\boldsymbol{x}')$ とすることで，判別器 r の確信度の合計を計算し，これを判別器の信頼度として扱う．

$$A_r = \sum_{\substack{\boldsymbol{x}'\subset\mathcal{X}'\\ c(\boldsymbol{x}')\in\mathcal{C}_r^+}} q_r(\boldsymbol{x}'). \quad (4)$$

確信度合計が大きい程，判別器が判別しやすい事を示し，逆に確信度が小さい程，判別しにくい事を示している．そこで判別器の精度に対応した，判別器毎の重みを決定する必要がある．確信度合計の最大値を 1 として基準化するために以下の式で判別器 r の重みを決定する．

$$w_r = \frac{A_r}{\max_{r'=1}^{R}(A_{r'})}. \quad (5)$$

4.2　判別器の精度のばらつきを考慮した GBT モデル

判別器毎の重みの導出法によって得られた重み w_r を GBT モデルに適用させることで，判別器の精度のばらつきを考慮した GBT モデルを構成する．3.3 節と同様，対数尤度関数 $F_{\mathrm{MGBT}}(\boldsymbol{p},\boldsymbol{x})$ は次式で定義される．

$$F_{\mathrm{MGBT}}(\boldsymbol{p},\boldsymbol{x}) = \sum_{r=1}^{R} w_r\Bigg(q_r(\boldsymbol{x})\ln\frac{\sum_{c_k\in\mathcal{C}_r^+} p_k}{\sum_{c_\ell\in\mathcal{C}_r} p_\ell} + (1-q_r(\boldsymbol{x}))\ln\frac{\sum_{c_k\in\mathcal{C}_r^-} p_k}{\sum_{c_\ell\in\mathcal{C}_r} p_\ell}\Bigg). \quad (6)$$

GBT モデルを用いた多値判別器と同様，提案手法でも $F_{\mathrm{MGBT}}(\boldsymbol{p},\boldsymbol{x})$ を最大とする $\hat{p}_k$ を $p(c_k|\boldsymbol{x})$ の推定値と見なす．

5　実験方法

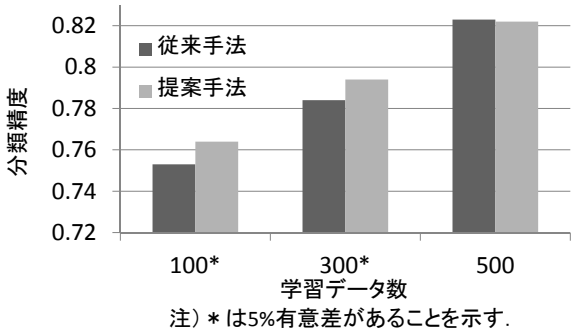
提案手法の有効性を検討するため，新聞記事を用いて分類実験を行い，分類精度の評価を行った．

5.1　実験条件

実験には，毎日新聞 2000 年の 4 カテゴリ (社会・スポーツ・芸能・経済) の記事を使用する．すべての記事は 1 カテゴリだけに属し，カテゴリの重複はない．学習データを 1 カテゴリ 100 件，300 件，500 件として，それぞれ 5 回繰り返し，その平均値によって評価を行う．テストデータは一律 200 件とする．対数尤度のパラメータ $\boldsymbol{p}$ の推定には勾配法を用いた．特徴量としては単語頻度を使い，文書頻度 10 以上の単語によって特徴量空間を構成する．比較手法は，判別器の精度を考慮しない従来の GBT モデルを用いた．

5.2　実験結果

学習データ数が 100 件，300 件，500 件の 3 パターンをそれぞれ 5 回ずつ実験し，分類精度の平均を求めた結果を図 1 に示す．結果より，学習データ数が 100 件，300 件の場合には，提案手法は従来手法よりも分類精度が有意に高く，500 件の場合は有意差がなかった．学習データ数が少ない場合の提案手法の有効性を示すことができた．



注）＊は5%有意差があることを示す．

図 1. 各学習データ数による分類精度

5.3　考察

提案手法は学習データ数が少ない場合は有効な手法といえる．学習データ数が多い場合は，提案手法の効果が薄れたが，これは各判別器の分類精度が向上し重み付けの効果が少なくなったからだと考えられる．

6　まとめと今後の課題

本研究では GBT モデルによる多値判別器の構成に関して，判別器の精度のばらつきを考慮した多値判別手法を提案し，実験によって，その有効性を示した．今後の課題として，最適な重みを導出する手法や，任意の判別器の組み合わせにおける重みを導出する手法の検討が挙げられる．

参考文献

[1] 大山賀己, 竹之内高志, 石井信, “ＥＣＯＣ復号法に基づく階層的多値判別法,” 電子情報通信学会技術研究報告, vol. 107, no. 542, NC 2007-169, pp. 337–342, 2008 年 3 月.

[2] Tzu-Kuo Huang, “Generalized bradley-terry models and multi-class probability estimates,” *The Journal of Machine Learning Research* 7, pp. 85–115, Jan. 2006.

[3] 池田思朗, “2 クラス判別器の組み合せによる多クラス判別 統計モデルとパラメータ推定,” 統計数理第 58 巻第 2 号, pp. 157–166, 2010 年 3 月.