Vol.2017-MPS-115 No.16
2017/9/26

FFDを用いた二値分類のための次元削減法に関する一考察

大窪 啓介[†1,a)]　雲居 玄道[†1,b)]　後藤 正幸[†1]

概要：高性能な非線形二値分類器のひとつに FFD(Fast Flux Discriminant) がある．この手法の特徴は，変数間の交互作用を考慮した分類モデルも構成できる点にある．そのため，実際に分類器の学習を行う前に，カーネル密度推定に基づいて特徴値の生成を行うことで，変数間の交互作用を考慮している．しかし，全変数間の交互作用を考慮するため，次元が大きくなると，計算量が指数的に増大することが欠点として挙げられる．そこで本研究では，FFD において，相互情報量を用いて事前にデータの次元を削減する方法を提案する．これにより，次元が大きいときにも交互作用を考慮できるようにし，分類精度を向上することを目的とする．さらに，新聞記事データの文書分類問題に適用し，その有効性を示す．

1. はじめに

近年，膨大かつ多種多様な電子化文書が氾濫しているため，文書の効率整理や，不要な情報のフィルタリング技術へのニーズが高まり，文書の自動分類の研究が盛んに行われている．多くの応用例を持つ二値分類に対しては，サポートベクトルマシンなどの様々な手法が適用可能である．二値分類器の中でも，高性能でかつ，大規模な非線形分類に用いられている手法の 1 つとして FFD(Fast Flux Discriminant)[1] がある．この手法が持つ他の二値分類器にはない利点として，線形モデルの効率性と非線形モデルの精度を同時に満たすことが挙げられる．

さらに，変数間の交互作用を考慮した分類モデルを構成できる点についても利点として挙げられる．しかし，全変数間の交互作用を考慮するため，次元が大きくなると，計算量が指数的に増大してしまう．そのため，多くの単語を用いる文書分類問題に対しては，変数間の交互作用を考慮した FFD は計算量的には実現できないという問題がある．

そこで本稿では，FFD において，事前にデータの次元を削減する方法を提案する．具体的には，カテゴリ間との相互情報量がある閾値よりも大きい単語を選択し，学習に用いるようにする．これにより，多くの単語を用いる文書分類問題に対しても，分類に寄与する変数間の交互作用を考慮できるようにすることで，分類精度の向上が期待できる．提案手法の有効性を検証するため，新聞記事の分類問題に適用し，分類精度などの観点から評価を行う．

†1 現在，早稲田大学
Presently with Waseda Uniersity. , Shinjuku, Tokyo 169–0072, Japan
a) tmc1374-ko@asagi.waseda.jp
b) moto-aries@ruri.waseda.jp

2. FFDとその問題点

データ数 N の学習データ集合を以下で表す．

$$Data = \{(X_i, Y_i) | X_i \in R^D, Y_i \in \{-1, 1\}\}_{i=1}^{N} \quad (1)$$

ただし，R^D は D 次元ユークリッド空間とする．FFD の学習は変数間の交互作用を考慮する処理と，それを受けて実際に分類学習を行う処理に別れている．そのため，FFD を実行する際に交互作用を考慮せず，分類学習のみを行うことも可能である．しかし，変数間の交互作用を考慮した分類モデルを構成しようとすると，全変数間の交互作用を考慮することになる．そのため，次元が数千から数万規模になると，計算量が指数的に増大してしまい，結果として FFD を実行できなくなってしまう．実際，次元数 D の 2 乗以上のオーダの計算が必要となる．そのため，多くの単語を用いる文書分類問題においては，変数間の交互作用の考慮ができないという問題がある．

3. 提案手法

3.1 相互情報量を用いた特徴選択

あるデータを $x_d (1 \leq d \leq D)$，カテゴリを $c_k \in C (1 \leq k \leq K)$ とする．この時，単語 x_d とカテゴリ間の相互情報量 $MI(x_d, C)$[2] を以下のように定義する．

定義 3.1 (相互情報量)

$$MI(x_d, C) = \sum_{k=1}^{K} P(x_d, c_k) \log \frac{P(x_d, c_k)}{P(x_d)P(c_k)} \quad (2)$$

$P(x_d, c_k)$： 全文書中で単語 x_d を含み，かつカテゴリ c_k に属する文書の割合

$P(x_d)$： 全文書中で単語 x_d を含む文書の割合
$P(c_k)$： 全文書中であるカテゴリ c_k に属する文書の割合

ここで，定義 1 において，相互情報量が大きな値をとる単語 (特徴) は，カテゴリ間でその単語の出現文書に偏りがあり，かつ出現文書数の等しい単語であるといえる．つまり，相互情報量の値が大きい単語のみを用いることで，計算量が削減できると共に，分類に寄与する変数と変数間の交互作用を考慮できると考えられる．

3.2 FFD における次元削減への適用

各次元におけるデータとカテゴリ間の相互情報量を求める．そして，求めた相互情報量を降順に並べ，その上位の単語のみを用いることで次元削減を行う．

ここで，相互情報量を求めるために必要なそれぞれの確率について見ていく．まず，本稿では二値分類を対象とするため，$K = 2$ となり，$P(c_k)$ は 0.5 となる．次に，$P(x_d)$ については，d 番目の次元における全てのデータ数 N 個のうち，0 より上の値を持つデータ数 a_d 個の割合とする．

$$P(x_d) = \frac{a_d}{N} \tag{3}$$

次に，$P(x_d, c_k)$ については，d 番目の次元における全てのデータ数 N 個のうち，0 より上の値を持ち，かつ y の値が -1 もしくは 1 であることを満たすデータ数をそれぞれ求め，それらの値と全体のデータ数との割合とする．

こうして求めたそれぞれの確率と (2) 式を元に各単語における相互情報量を求める．そして，求めた相互情報量がある閾値よりも大きい単語を実際の分類学習に用いる．

なお，この閾値が小さすぎると，次元削減が少なく，交互作用が考慮できなくなってしまう．そのため，以下の実験は，この閾値の最適な値を探索的に求めた上で行った．

4. 実験

本稿では，2015 年度の読売新聞のデータを用いて実験を行った．この新聞記事データは，政治，経済，スポーツ，社会，文化，生活，犯罪事件，科学の 8 個のカテゴリに分けられ，詳細については表 1 に示す．本稿では，比較手法として交互作用を考慮せず，全単語を用いた FFD を用い，提案手法については，学習の際に交互作用を考慮しなかった場合と考慮した場合の 3 通りの実験を行った．また，実験は FFD を用いて 1 対他の二値分類 [3] を行った．

なお，本実験では相互情報量の閾値は，事前の分析から 4.0×10^{-3} に設定した．それぞれの分類精度，及び提案手法における次元数は表 2 のようになった．

なお，表 2 におけるカテゴリ番号とは，1 対他分類においてその番号のカテゴリとそれ以外の複数のカテゴリとで二値分類を行ったことを表す．例えば，カテゴリ番号 1 については，カテゴリ 1 とそれ以外とで分類を行っている．

表 2 より，交互作用を考慮しない提案手法は，比較手法と比べて精度がほとんど変わらないことが分かる．このことから，提案手法は精度を維持しつつ次元を削減できていると考えられる．この理由としては，次元削減された相互情報量が小さい単語は文書データにおける出現頻度が 0 もしくは 0 に近いものが多く，分類学習にあまり影響していないためだと考えられる．また，提案手法において交互作用を考慮した時の方が考慮していない時よりも精度が上回っていることが分かる．このことから，提案手法によってより精度の高い分類器が作成できたと考えられる．また，この理由としては，提案手法によって分類に寄与する変数間の交互作用を考慮できたことで，各次元における細かい分類が可能になったためだと考えられる．

表 1 データの概要 (2015 年 読売新聞)

文書の特徴ベクトル (次元)	40067
データ数	合計 12000 件
訓練データ	1350 件/カテゴリ，合計 10800 件
テストデータ	150 件/カテゴリ，合計 1200 件

表 2 1 対他分類における各分類精度

カテゴリ	従来手法	提案手法 (次元数)	
番号	交互作用無	交互作用無	交互作用有
1	86.08%	86.00%(2265)	91.25%(2265)
2	86.00%	86.00%(2445)	95.50%(2445)
3	85.83%	85.33%(1895)	88.25%(1895)
4	85.67%	85.67%(1683)	90.58%(1683)
5	85.33%	85.25%(1886)	90.58%(1886)
6	84.58%	83.25%(1444)	88.78%(1444)
7	85.58%	85.33%(1914)	89.75%(1914)
8	86.83%	86.25%(2007)	90.09%(2007)

5. まとめと今後の課題

本稿では，FFD において，分類に寄与する変数間の交互作用を考慮できるようにするために，相互情報量を用いて事前にデータの次元を削減する方法を示した．その結果，次元が削減され，交互作用を考慮することで分類精度が向上する結果となった．文書分類のように，特徴空間の次元が大規模になる問題に対して，提案手法は有効と考えられる．また，今後の課題としては，文書データ以外の分類問題に適用することで，本手法における交互作用の有効性を検討していくことが挙げられる．

参考文献

[1] Wenlin Chen, Yixin Chen, Kilian Q. Weinberger: *Fast Flux Discriminant for Large-Scale Sparse Nonlinear Classification*, Proc.ACM SIGKDD Coference (KDD-14).
[2] 津田裕一，山岸英貴，石田崇，平澤茂一：相互情報量に基づく特徴選択を用いた文書自動分類，第 4 回情報科学技術フォーラム (2005).
[3] 後藤正幸，小林学：入門 パターン認識と機械学習，コロナ社 (2014).