LG-5 順序関係を推定する順位づけ学習問題とその解法

Rank Estimation : A New Challenge to Machine Learning

賀沢 秀人† 平尾 努† 前田 英作†
Hideto Kazawa Tsutomu Hirao Eisaku Maeda

1 はじめに

本稿では、我々が提案する新しい機械学習の問題である「順位づけ学習問題」を取り上げる [4]。順位づけ学習問題とは、ある順序を持つ要素の集合があるときに、有限個のサンプルとその順位から、集合全体での順序を推定する問題である。

順位づけ学習問題の例として、競馬の着順予測がある。競馬の着順予測では、まず、過去のレースに出走した馬について、体高・体重などの定量的なデータと各馬の順位が与えられる。そして将来のレースについて、各出走馬のデータからそれらの順位を予測することが目的となる。このとき、予測するレースにおける出走馬は過去のレースに出走している必要性はなく、全く未知の馬である可能性もある点が重要である。

また、人間による総合的な判断を学習する場合にも、同様な状況がしばしば発生する。例えば、文書検索では、検索要求に適合する文書をその「適合度」順に回答することが求められるが、そのときサンプルとして与えられるのは各文書の適合順位のみであり、定量的な「適合度」といったものは与えられない。

このように、現実の問題では、サンプルとして要素間の相対的な順序しかわからず、その順序のもととなる「スコア」のようなものは観測できないことが多い。そのため順位づけ学習問題に対して有効な解法を用意することは、機械学習の適用範囲を拡大する意味で非常に重要である。

しかし、既存の機械学習手法では順位づけ学習問題に対して満足な解は与えられない。これは、従来の機械学習の枠組みでは、サンプルとして各要素ごとに絶対的な値が与えられることが仮定されているためである。[1]

本稿では、まず、順位づけ学習問題を数学的に定式化したのち、それに対する幾つかの解法を説明する。次に、人工データと現実の新聞記事から作成した重要文抽出データ [5] を用い、それらの解法を比較した結果について報告する。

2 順位づけ学習問題の定式化

順位づけ学習問題を次のように定式化する。

全順序 $\succcurlyeq$ が定義された集合 X と、その上の確率分布 $p(x)$ が与えられたとする。さらに、関数 $o(x,x')(x,x' \in X)$ を、$x \succcurlyeq x'$ のとき 1、それ以外のとき -1 を取る関数とする。

いま、サンプルとして $p(x)$ にしたがって得られた要素 $x_1, x_2, \ldots, x_m$、と $\succcurlyeq$ にもとづいた順位づけ結果が与えられたときに、次の $R[g]$ を最小にする関数 $g(x)$ を求める。

$$R[g] = \int \theta[-(g(x) - g(x'))o(x,x')]p(x)p(x')dxdx' \quad (1)$$

ただし $\theta(z) = 1(z \geq 0), -1(z < 0)$ とする。

$o(x,x')$ の定義から、x, x' について $\succcurlyeq$ が与える順序と、$g(x), g(x')$ の大小関係が一致したときに $\theta[-(g(x) - g(x'))o(x,x')]$ は 0、一致しないときに 1 となる。したがって、$R[g]$ は $\succcurlyeq$ によって与えられる順序と、$g(x)$ の大小関係によって与えられる順序の、平均的な不一致度に相当する。

3 順位づけ学習問題の解法

前節で定式化した順位づけ学習問題の解法として、本節では support vector machine (SVM) をもとにした手法を説明する。これは SVM の利点として、次のような点があるためである。

- 高次元のデータを効率良く扱うことが出来る。SVM では学習の計算時間がデータの次元数と無関係なため、高次元のデータの扱いに適している。
- 学習結果が一意に定まる。SVM では学習問題の解が一意に定まることが知られており、ニューラルネットワークなどのような局所解の問題がない。

3.1 Support Vector Ordinal Regression

式 (1) において、(x,x') の対を一つの事例、$o(x,x')(= \pm 1)$ をそのカテゴリとみなすと、$X \times X$ 上の二値分類を学習する問題と考えることが出来る。したがって、既存の二値分類の学習手法との類推で、順位づけ学習問題を解くことができる。

Herbrich らは SVM をもとにした次のような手法を提案している [3][2]。そこでは、以下の目的関数を最小化することで $g(x) = w \cdot x$ を求める。

$$\min_{w,\xi} \frac{1}{2}|w|^2 + C\sum_{i=1}^{m}\sum_{j=1}^{m}\xi_{ij} \quad (2)$$

$$s.t. \quad o(x_i,x_j)(w \cdot x_i - w \cdot x_j) \geq 1 - \xi_{ij}, \xi_{ij} \geq 0 \quad (3)$$

式 (3) は、$g(x_i), g(x_j)$ の大小関係が $o(x_i,x_j)$ と一致するという条件を示し、式 (2) は、その条件が破られたときのペナルティを、過学習を避けるために $g(x)$ が滑らかな関数となるような正規化項とともに、最小化することを示している。以下、この手法を support vector ordinal regression (SVOR) と呼ぶ。

3.2 Ranking SVM

前節で説明した SVOR は、サンプル間の関係を直接的に利用した手法である。それに対し筆者らは、あるサンプルが特定の順位以上に出現する確率 (累積確率) の大小関係が、順序 $\succcurlyeq$ と一致することにもとづき、SVM により累積確率の推定を行うことで順位づけ学習を行う手法を提案している [4]。

ある x を含む n 個のサンプルを取り出したときに、x が第 s 位以上に出現する累積確率を $q_{s,n}(x)$ とする。$q_{s,n}(x)$ の推定は、m 個のサンプルを取り出し、その中で s 位以上で出現したサンプルを正例、それ以外を負例とした 2 カテゴリの分布関数の推定問題と考えることが出来る。

平尾らは、SVM を用いてある単独の s について $q_{s,n}(x)$ の推定を行い、それを $g(x)$ とすることで順位づけを行った [1]。[3] そ

---

† 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

[1] 例えば、回帰関数の推定では、サンプルとして各要素毎にある関数の値が与えられたときに、その関数を推定することが目的となっている。

[2] [3] で取り上げられている問題は順序づきカテゴリ間の分類問題であり、順位づけ学習問題ではない。

[3] これは、SVM がある近似のもとで (条件つき) 確率分布の推定を行っ

れに対し、筆者らはSVMを用いて複数のsについての$q_{s,n}(x)$の推定を行い、それを平均化したものを$g(x)$とすることで精度を向上させることを提案している[4]。

ここで、複数のsについての$q_{s,n}(x)$の推定と平均化は、以下の最適化を行うことで同時に行われる。(Ranking SVM)

■Ranking Support Vector Machine

■学習　n個1組の順位づきデータm組$x_i^1 \succcurlyeq x_i^2 \succcurlyeq \ldots \succcurlyeq x_i^n (1 \leq i \leq m)$が与えられたとき、以下の最小化を行う。

$$\min_{w,v_s,b_s,\xi_{is}^j} \frac{1}{2}|w|^2 + \frac{\lambda}{2}\sum_{s=1}^{n-1}|v_s|^2 + C\sum_{s=1}^{n-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\xi_{is}^j \quad (4)$$

$$(w+v_s)\cdot x_i^j \geq b_s + 1 - \xi_{is}^j \quad if\ j \leq s \quad (5)$$

$$(w+v_s)\cdot x_i^j \leq b_s - 1 + \xi_{is}^j \quad if\ j > s \quad (6)$$

$$\xi_{is}^j \geq 0 \quad (7)$$

■順位づけ（適用）　各入力xに対し、以下の$g_{rank}(x)$でソートし順位づけを行う。

$$g_{rank}(x) = w \cdot x \quad (8)$$

Ranking SVMでは、全ての順位についての境界面を一度に構成しつつ(式(5)(6))、wがそれらの境界面の方向ベクトルの平均になるように、各境界面ごとのばらつき(v^s)が小さくなるような最適化を行う(式(4))。ばらつきをどの程度許容するかはパラメータλでコントロールできる。

以上の最適化問題は次の双対問題と同値である。

$$\max_{\alpha_{i,s}^j} \sum_s\sum_{ij}\alpha_{i,s}^j - \frac{1}{2}\sum_{s,s'}\sum_{i,j,i',j'}\left(1+\frac{\delta_{s,s'}}{\lambda}\right) \times \alpha_{i,s}^j \alpha_{i',s'}^{j'} r(j,s) r(j',s')(x_i^j \cdot x_{i'}^{j'}) \quad (9)$$

$$\sum_{ij}\alpha_{i,s}^j r(j,s) = 0, \quad 0 \leq \alpha_{i,s}^j \leq C \quad (10)$$

ただし、$r(j,s) = 1(j \leq s), -1(j > s)$である。また式(8)は$\alpha_{i,s}^j$を用いて次のように書ける。[4]

$$g_{rank}(x) = \sum_s\sum_{ij}\alpha_{i,s}^j r(j,s)(x_i^j \cdot x) \quad (11)$$

4 実験

以下の実験では、人工データとTSCの重要文抽出データを用いて、単独SVM[1]、SVOR[3]、RankingSVMの比較を行う。なお、SVMの実装にはTinySVM[5]を用い、SVORとRanking SVMはTinySVMをもとに独自に実装したものを使用した。

4.1 人工データ

人工データとして、$[0,1]^2$上の一様分布からサンプリングしたサンプル(x_1, x_2)を$(x_1 - 0.5)(x_2 - 0.5)$の大きさで順位づけしたデータを作成した。訓練データとしては、4個1組を10～200組作成し、評価用のテストデータとして、同じ手順により作成した2000組のデータを用意した。

図1に実験結果を示す。Ranking、SVORはそれぞれRanking SVMとSVORの結果、SVMは1位を正例、2～4位を負例とし学習を行ったSVMの結果である。カーネルは2次の多項式カーネルを用い、hyper parameterは$C = 1, \lambda = 100$とした。評価は4個1組のテストデータを、それぞれの手法で学習した$g(x)$の大きさ順に並び替え、1位になったものが正しいか否かの正解率で評価した。Ranking SVM、SVORともにSVMよりも高い精度を示していることがわかる。

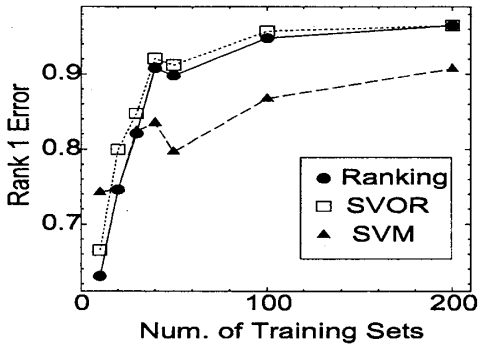

図1: 人工データによる比較(1位の正解率)

手法 ＼ 評価順位	10%	30%	50%
Ranking SVM	**0.60**	**0.54**	0.66
SVOR	0.56	0.50	0.64
SVM(50%)	0.42	0.47	**0.67**

表1: 重要文データによる比較

4.2 重要文抽出データ

新聞記事から重要度順に、全文数の10・30・50%にあたる文を選択したデータ(TSCの重要文データ)を用いて実験を行った[5]。TSCデータでは全てのサンプル(文)に順位が付与されているわけではないため、これまでの問題設定とは必ずしも一致しない。したがって、以下の実験では、(Ranking)SVMでは上位下位がわかる順位のみ対象として境界面を設定し、SVORでは順序関係がわかるサンプル対のみ対象として学習を行った。

実験の対象としては「社会」のジャンルにあたる63記事を用いた。各文は、記事中の位置・頻出単語の密度・機能語の種類などからなる約500個の二値素性のベクトルに変換した。[6]

学習にあたっては2次の多項式カーネルを用い、$C = 0.001, \lambda = 100$とした。また評価にあたっては63記事を訓練60・テスト3の割合で分けたものを用いて21-foldで、上位10・30・50%の正解率を計算した。

表1に実験結果を示す。SVM(50%)は重要度で上位50%に入る文を正例、それ以外を負例としてSVMで学習した結果である。また左列から順に10,30,50%までの正解率で評価した結果である。人工データの場合と同様、Ranking SVMとSVORで、平均してSVMよりも高い精度が得られた。また、その精度でもRanking SVMは高い精度を達成し、SVMのような極端な精度劣化は見られなかった。

[1] 平尾他. Support Vector Machineによる重要文抽出, 情処FI研(2001/7)

[2] Herbrich. Learning Kernel Classifiers, The MIT Press (2002)

[3] Herbrich他. *Large Margin Rank Boundaries for Ordinal Regression*, Advances in Large Margin Classifiers, The MIT Press (2000)

[4] 賀沢他. Ranking SVMによる順位づけ学習と重要文抽出への応用, IBIS2002投稿中.

[5] Fukushima他. Text Summarizataion Challenge, Proc. of NAACL2001 Workshop on Automatic summarization (2001)

ているとみなせることに基づく。[2]

[4] 式(9)(11)はxに関して内積計算しか必要ないので、通常のSVMと同様kernelに置き換えることが可能である。

[5] http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/

[6] データの前処理に関しては[1]を参照のこと。