



ストリームデータ環境下での 組み合わせ素性を用いた 効率的な予測器・分類器学習法

東京大学 情報理工学系研究科

大岩秀和 中川裕志



概要

- 組み合わせ素性を用いた学習・予測

- 例：マイクロブログ



- 1データ当たりの情報量が非常に少

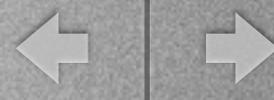
- より深い学習・予測を可能にしたい

上海+香港、上海+パリetc...



問題点

- 全組み合わせ素性の利用は非現実的
 - メモリ容量 : $O(2^N)$
 - 計算量 : $O(2^N)$ $N \dots$ 素性数
- 利用する組み合わせ素性を事前に指定
 - コンセプトドリフトへの対応が困難
- カーネルを利用
 - 各組み合わせ素性の予測寄与度を評価するのが困難



目標

- 組み合わせ素性を用いた**効率的な学習**
- 利用する組み合わせ素性を予め**指定しない**
- **リミテッドメモリ**での動作を可能に
- 組み合わせ素性の予測寄与を陽に示す
- 全データが一度に展開不可能な**ストリームデータ環境**下での学習・予測を可能に



組み合わせ素性

- 素性の組み合わせで表現される素性
- N-gram素性も組み合わせ素性的一种
- 本研究では、BOW表現の各要素の組み合わせを組み合わせ素性として扱う



既存手法 [1/2]

- 分割多項式カーネル **Passive-Aggressive** [Yoshinaga+, 10]
 - 頻出する組み合わせ素性は陽に表現
 - 線形予測器として学習
 - その他の組み合わせ素性は多項式カーネルで学習
 - 線形予測器とカーネル法の良いところ取り
 - さらなる高速化手段として、**Kernel Slicing**を提案



既存手法 [2/2]

- Forgetron [Dekel+, 08]
 - 基本は、パーセプトロン + カーネル
 - 不必要なカーネルを忘却し、メモリ容量に上限を課す
 - Mistake Boundを証明
- LIロジスティック回帰 + Grafting [Okanojara and Tsujii, 09]
 - Grafting : 対数尤度を最も減少させる素性を逐次的に追加
 - LI正則化により、最適解におけるパラメータ数を減少



用いる学習手法

- オンライン学習 + LI正則化 → FOBOS

オンライン学習

一度に一データのみを
使用し、学習を行う

全組み合わせ素性を

一度に処理する必要がなくなる

LI正則化

有効でない素性を零化
動的な特徴選択

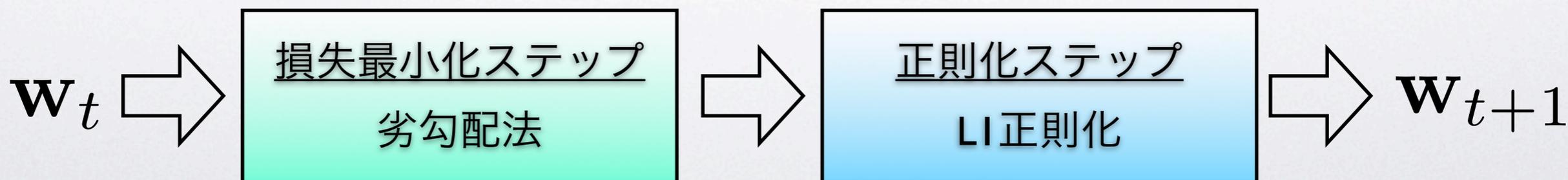
学習しながら、予測に有用な
組み合わせ素性のみを残す



Forward Backward Splitting

[Duchi+, 09]

- 更新ステップを2段階に分割
 - 損失最小化ステップ
 - 正則化ステップ
- 分類問題・回帰問題等に適用可能
- Regretによる収束性の証明もなされている
- 詳しくは、こちらの[PDF](#)参照





FOBOSを用いる利点

- 全組み合わせ素性の情報を同時に扱う必要がない
 - 不要な組み合わせ素性は逐次的に零化
 - 学習・予測時に用いる組み合わせ素性は、データ中に現れたもののみ
 - したがって、リミテッドメモリ環境下でも動作可能
- 全組み合わせ素性を陽に扱える
 - 0-1以外の値をとる特徴
 - コンセプトドリフトにも動的に対応可能



擬似コード

Algorithm 1 Conjunctive Features Forward Backward Splitting (CFFOBOS)

Input: $\mathbf{w}_1=\mathbf{0}$ (weight vector), $\mathbf{cw}_1=\mathbf{0}$ (weight vector for conjunctive features)

for $t = 1$ to T **do**

Receive a feature vector \mathbf{fv}_t and a label y_t

$\mathbf{cfv}_t = \text{CreateCFV}(\mathbf{fv}_t)$; // Create conjunctive feature

$\text{score}_t = \mathbf{fv}_t * \mathbf{w}_t + \mathbf{cfv}_t * \mathbf{cw}_t$; // Predict label and calculate margin

if $\text{score}_t \cdot y_t \leq 1$ **then**

$(\mathbf{w}_{t+\frac{1}{2}}, \mathbf{cw}_{t+\frac{1}{2}}) = \text{Update}(\mathbf{fv}_t, \mathbf{cfv}_t, y_t, y_t \cdot \text{score}_t)$; // Subgradient Method

else if **then**

$(\mathbf{w}_{t+\frac{1}{2}}, \mathbf{cw}_{t+\frac{1}{2}}) = (\mathbf{w}_t, \mathbf{cw}_t)$

end if

$(\mathbf{w}_{t+1}, \mathbf{cw}_{t+1}) = \text{Truncation}(\mathbf{w}_{t+\frac{1}{2}}, \mathbf{cw}_{t+\frac{1}{2}})$; //Apply L1-regularization

end for

Output: $\mathbf{w}_{T+1}, \mathbf{cw}_{T+1}$

損失最小化ステップ

L1正則化ステップ



実験

- 2値分類問題
 - 組み合わせ素性を用いたFOBOSと用いないFOBOSで精度を比較
 - 組み合わせ素性は2次のもののみを使用
 - 損失関数はHinge-Loss $l(\mathbf{w}) = [1 - y\mathbf{w} \cdot \mathbf{x}]_+$
 - データセットの反復回数は10回
 - L1正則化項のハイパーパラメータを様々な値に変化



Amazon 評価分類データセット

[Blitzer+, 07]

- 特定のトピックに関するレビューのポジティブ・ネガティブ判定

	訓練数	テスト数	素性数	組み合わせ 素性数
books	4,019	446	310,266	48,132,340,245
dvd	3,228	358	264,054	34,862,125,431
electronics	5,113	568	218,474	23,865,335,101
kitchen	5,351	594	190,991	18,238,685,545



実験結果

4データセットの平均精度

L1正則化項

$$r(\mathbf{w}) = \lambda \|\mathbf{w}\|_1$$

λ	CFFOBOS	FOBOS
0.2	53.93	53.93
0.1	62.47	62.15
0.05	67.96	67.90
0.02	75.15	74.48
0.01	78.36	79.37
0.005	82.97	83.71
0.002	85.48	85.02
0.001	85.02	85.84

立ち上がりの
精度は良い

booksで4-6%精度
が落ちている

CFFOBOSは
過学習しやすい



ニュース記事分類

[Lewis+, 04]

	訓練数	テスト数	素性数	組み合わせ 素性数
rcv-1	20,242	677,399	47,237	1,115,643,466

	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
CFFOBOS	0.14 (0.14)	14.44 (14.44)	41.29 (41.30)	53.99 (54.01)	63.92 (65.45)	74.00 (75.17)	89.11 (89.25)	92.34 (92.05)
FOBOS	0.14 (0.14)	14.43 (14.43)	41.31 (41.31)	53.01 (53.03)	63.71 (65.45)	73.43 (72.76)	88.88 (87.68)	91.98 (91.64)

()内は、1回反復した時の精度



参考文献リスト

- The Forgetron: A Kernel-Based Perceptron on a Budget. [Dekel, Shalev-shwartz, and Singer, 2008]
- Efficient Online and Batch Learning using Forward Backward Splitting. [Duchi and Singer, 2009]
- Learning Combination Features with L1 Regularization. [Okanohara and Tsujii, 2009]
- Kernel Slicing : Scalable Online Training with Conjunctive Features. [Yoshinaga and Kitsuregawa, 2010]
- Polynomial to Linear: Efficient Classification with Conjunctive Features. [Yoshinaga and Kitsuregawa, 2009]
- 素性の組み合わせを実現する Power Set Kernel とその高速化. [工藤 and 松本, 2003]
- Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. [Blitzer, Dredze, and Pereira, 2007]
- RCV1: A new benchmark collection for text categorization research. [Lewis, Yang, Rose, and Li, 2004]