

自然言語処理研究会@NAIST 11/21/2000

Support Vector Machine を用いた Chunk 同定

奈良先端大情報科学研究科 自然言語処理学講座

工藤 拓, 松本 裕治

{taku-ku, matsu}@is.aist-nara.ac.jp

Chunk 同定問題 (Chunking)

- 文 (相当のもの) をある基準で分割し, まとめる (同定)
- まとめた各要素にタグを付ける (分類)



NLP の多くの問題は, Chunk 同定問題としてみなすことができる (日本語わかち書き, 英語 tokenization, 形態素解析, 文節切り, baseNP 抽出, Named Entity 抽出, 専門用語抽出...)

マージン最大化戦略に基づく学習器 (Large Margin Classifiers) の一種である Support Vector Machine , 重み付き多数決 を用いて, 一般的な Chunk 同定問題を扱う

Outline

- Support Vector Machines (マージン最大化, 汎化能力)
- 重み付き多数決
- Chunk 同定問題のモデル (Inside/Outside)
- SVM/重み付き多数決 による Chunk 同定
- 実験結果, 考察 (baseNP Chunking, Chunking)
- 関連研究との比較
- 今後の課題
- まとめ

Support Vector Machines

- 線形, 2 値 (正例/負例) 分類器 ($w \cdot x + b = 0$)
- マージン最大化に基づき最適な分離平面 (w, b) を決定

- Support Vector 以外のデータは結果に影響を及ぼさない
- 線形分離不可能の時も, Kernel 関数 を用いるとことで対処

SVM の汎化能力 (1/4)

VC Bound

$$E_g[f] \leq \underline{E_t[f]} + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l} \right)}$$

↓

訓練データに対するエラー率

Empirical Risk

↓

モデルのキャパシティ

VC confidence

l : 学習データ数

h : *VC dimension*(モデルの複雑さを示すパラメータ)

SVM の汎化能力 (2/4)

SVM の汎化能力 (3/4)

事例の次元数を n , マージンを M , 全事例を囲む球面の最小直径を D

$$h \leq \min(D^2/M^2, n) + 1$$

- n が十分大きいならば h の上限は, 事例の次元数 (素性数) に依存しない
- マージン M 最大化 $\rightarrow h$ 最小
- D は Kernel 関数に依存,
VC dimension (h) は Kernel 関数選択の基準を与える

SVM の 汎化能力 (4/4)

$E_l[f]$ を *Leave-One-Out* によって評価されるエラー率とする場合

$$E_l[f] \leq \frac{\text{Number of Support Vectors}}{\text{Number of training samples}}$$

Leave-One-Out: l 個の学習データのうち 1 個をとりのぞいてテストデータとし, 残り $l-1$ を使って学習することをすべてのデータについて l 回くりかえす

- SV のみ が 分離平面 を 決定する という 事実 から 容易に 導出可能
- SV の数 が 少ない ほど 汎化能力 が 高い (オッカムの剃刀, MDL)

重み付き多数決

仮説 (学習アルゴリズム) : H_1, H_2, \dots, H_T

重み : $\alpha_1, \alpha_2, \dots, \alpha_T$

$$y = \text{sgn}\left\{\sum_{i=1}^T \alpha_i H_i(\mathbf{x})\right\}$$

異なる視点で、異なる素性に過学習した仮説の線形結合



最終的な分離平面は個々の分離平面の中間に設置される



マージン最大化の効果 [Schapire 96]

Chunk の表現方法 (1/2)

Inside/Outside [Ramshaw, Marcus 95, Tjong Kim Sang 2000]

- Chunk の中 (I), 外 (O) を基本単位 + それ以外のタグ (B/E)
- IOB1/IOB2/IOE1/IOE2

Chunk の表現方法 (2/2)

	Base NP Chunking				Chunking
	IOB1	IOB2	IOE1	IOE2	IOB2
In	O	O	O	O	B-PP
early	I	B	I	I	B-NP
trading	I	I	I	E	I-NP
in	O	O	O	O	B-PP
busy	I	B	I	I	B-NP
Hong	I	I	I	I	I-NP
Kong	I	I	E	E	I-NP
Monday	B	B	I	E	B-NP
,	O	O	O	O	O
gold	I	B	I	E	B-NP
was	O	O	O	O	B-VP

SVM による Chunk 同定 (1/2)

- Inside/Outside を使うことで, Chunk 同定問題をタグ付与問題に帰着
- 従来からあるタグ付与問題 (POS tagging など) の技術が容易に応用可能
- タグの周辺のコンテキストを素性とみなし, 現在のタグを精度よく推定するルールを導く処理 → 機械学習の応用



- SVM と 重み付き多数決により Chunk 同定 (= タグ付与) を行なう
- 2 値分類器から多値分類器へ拡張 → *pairwise method* を使用

SVM による Chunk 同定 (素性) (2/2)

- Chunk 素性は左から右に解析する課程で動的に追加 (前向き解析)
- 右 2 つの Chunk を素性として使い, 右から左に解析 (後ろ向き解析)

重み付き多数決による Chunk 同定 (1/2)

- $\{\text{IOB1}/\text{IOB2}/\text{IOE1}/\text{IOE2}\} \times \{\text{前向き}, \text{後ろ向き}\}$ 合計 8 種類の表現方法を用いて, SVM で独立に学習
- 個々の仮説の重み付き多数決を用いて最終的な Chunk を決定
- 重みの与え方が問題となる \rightarrow 4 種類の重みの決定方法を提案
 - 均一 (baseline)
 - 交差検定 (5-fold)
 - VC bound
 - Leave-One-Out bound

重み付き多数決による Chunk 同定 (2/2)

- VC bound

$$h \leq \min(D^2/M^2, n) + 1$$

$$E_g[f] \leq E_t[f] + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}\right)}$$

- Leave-One-Out bound

$$E_l[f] \leq \frac{\text{Number of Support Vectors}}{\text{Number of training samples}}$$

これらの bound が真のテストデータに対するエラー率を精度よく推定す能力を持っているなら、良い重みになるのではないか？

実験

3種類のデータセット

- **baseNP-S**: Penn Tree Bank/WSJ 15-18 を Training, 20 を Test, Brill Tagger で POS tagging, baseNP chunking に用いられる標準データとして広く認知
- **baseNP-L**: Penn Tree Bank/WSJ 02-21 を Training, 00 を Test, (ただし, 実験時間から, IOB1/IOE1 の表現方法, 交差検定による重み付けは行なっていない)
- **chunking**: CoNLL-2000 Shared Task に使用されたデータ. Penn Tree Bank/WSJ 15-18 を Training, 20 を Test VP, PP, ADJP, ADVP, CONJP, INITJ, LST, PRT, SBAR の合計 10 種類の英語の基本句が付与

$F_{\beta=1}$ 値 (再現率と適合率の調和平均) による評価

実験結果 (baseNP-S 個々の表現)

変換先	$F_{\beta=1}$	B	C	D
IOB1-前	93.76	.9394	.4310	.9193
IOB1-後	93.93	.9422	.4351	.9184
IOB2-前	93.84	.9410	.4415	.9172
IOB2-後	93.70	.9407	.4300	.9166
IOE1-前	93.73	.9386	.4274	.9183
IOE1-後	93.98	.9425	.4400	.9217
IOE2-前	93.98	.9409	.4350	.9180
IOE2-後	94.11	.9426	.4510	.9193

B:交差検定 C:VC bound D:Leave-One-Out bound

実験結果 (baseNP-L 個々の表現)

変換先	$F_{\beta=1}$	B	C	D
IOB2-前	95.34	-	.4500	.9497
IOB2-後	95.28	-	.4362	.9487
IOE2-前	95.32	-	.4467	.9496
IOE2-後	95.29	-	.4556	.9503

B:交差検定 C:VC bound D:Leave-One-Out bound

実験結果 (chunking, 個々の表現)

変換先	$F_{\beta=1}$	B	C	D
IOB1-前	93.48	.9342	.6585	.9605
IOB1-後	93.74	.9346	.6614	.9596
IOB2-前	93.46	.9341	.6809	.9586
IOB2-後	93.47	.9355	.6722	.9594
IOE1-前	93.45	.9335	.6533	.9589
IOE1-後	93.72	.9358	.6669	.9611
IOE2-前	93.45	.9341	.6740	.9606
IOE2-後	93.85	.9361	.6913	.9597

B:交差検定 C:VC bound D:Leave-One-Out bound

実験結果 (重み付き多数決)

学習データ	A	B	C	D
baseNP-S	94.16	94.22	94.22	94.18
baseNP-L	95.77	-	95.66	95.66
chunking	93.77	93.89	93.91	93.85

A:均一 B:交差検定 C:VC bound D:Leave-One-Out bound

考察

- 重み付き多数決の採用で、個々のどのれよりも高い精度を示す
- baseNP-S, chunking の手法に重み付け手法において、baseline よりも精度が向上
- VC-Bound は交差検定に匹敵する精度を示す
 - 交差検定は、分割個数が多くなると多くの計算量が必要
 - VC bound は一回の学習で高精度の重みが推定可能, 高効率
 - VC bound は複数のモデルから最良の物を選択する良い基準を与える.

VC bound の予測能力

Kernel 関数 (polynomial kernel) の次元数と VC bound の関係
(d 次の polynomial: d 個までの素性の組合わせを考慮)

Kernel 関数の次元数	1	2	3	4
VC bound	0.766	0.55	0.51	0.53
テストデータに対する F 値	91.98	93.84	93.69	93.18

- 素性の組合わせを考慮する必要性
- VC bound は Kernel 関数の次元数 (パラメータ) を選択する良い基準を与える

関連研究との比較

	手法の概要	精度
Tjong Kim Sang 2000	MBL, ME, IGTre e 等の 7 種類のア ルゴリズム, IOB1/IOB2/IOE1/IOE2 の 4 種 類の表現の重み付き多数決	baseNP-S 93.86 baseNP-L 94.22
工藤, 松本 2000	SVM + IOB2 の単独表現 CoNLL2000 Shared Task 中, 最良の結果	chunking 93.48
提案手法		baseNP-S 94.22 baseNP-L 95.77 chunking 93.91

- 単純な精度比較だと提案手法が優位
- SVM は従来からある機械学習アルゴリズム (MBL, ME, IGTre e) のどれよりも高性能

今後の課題

- 他の分野への応用, 検証
(形態素解析, わかち書き, 専門用語抽出, Named Entity 抽出)
- より予測能力の高い Bound を重みとして採用
(Span SVM [Chapelle, Vapnik 2000])
- 固定長モデル から 可変長モデル へ

可変長モデル (2/3)

2つのアプローチ

- Kernel 関数の optimize, 可変長もふくめた事例間の距離の定義
- Virtual SVM [Schölkopf, Buregs, Vapnik 1998]
 - タスク固有の不変性を機械学習に取り入れる手法
 - 固定長でまず学習,
 - 変換規則を用いて Support Example を展開, 学習データ (Virtual Training Example) として追加, この時, context が可変長になる
 - 再学習

まとめ

- Support Vector Machine (SVM) に基づく一般的な chunk 同定問題の解析手法を提案
- 過去の MBL や ME に基づくモデルよりも高い精度, SVM の持つ高い汎化能力を裏づける結果
- タグの表現方法が異なる複数のモデルの重み付き多数決を行なう事で, 個々のどのモデルよりも高い精度を示した
- 重みとして SVM に固有の手法 (VC bound, Leave-One-Out bound) を提示し, その有効性を検証