

模倣学習による決定的解析での誤り伝播の回避

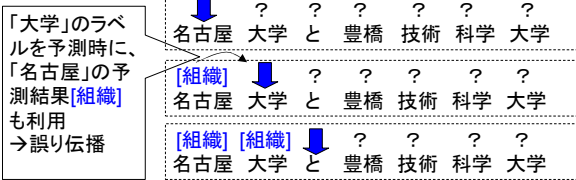
坪井祐太
日本アイ・ビー・エム株式会社東京基礎研究所
yutat@jp.ibm.com

概要

- 決定的解析とは
- 模倣学習とは
- DAGGERアルゴリズム
- 系列ラベリング問題への適用
 - 固有表現抽出(CoNLL2002)での実験結果
- 依存構造解析への適用
 - 英語依存構造解析での実験結果
- まとめと今後の課題

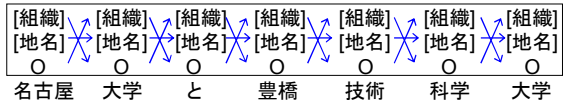
(本研究が対象とする)決定的解析とは

- 貪欲に予測(分類)を繰り返す解析手法
 - 系列ラベリングや依存構造解析など
- 固有表現抽出(系列ラベリング問題)の例



既存手法: 大域学習(Global Training)による誤り伝播の回避

- CRF, 構造化SVM, 構造化パーセプトロンなど
- 構造全体で最適な解を選ぶ

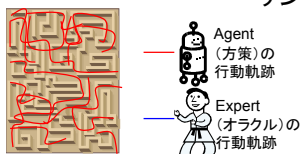


→ 指数個の解析候補を扱うために近似

1. 動的計画法: 解析候補から最適解を選択できる一方、特徴量を小範囲に限定
2. ビーム探索・リランキング: 広範囲の特徴量を使用する一方、解析候補から近似解を選択

模倣学習(Imitation Learning)とは

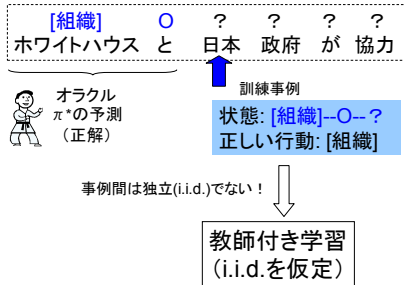
- 強化学習: 行動の系列と結果(報酬)のサンプルから学習(自ら動きながら学習)
- 模倣学習: エキスパートの行動列を参照し、報酬の多い行動・状態を効率的にサンプリング



本研究の貢献

- NLPタスクへの模倣学習の適用事例
 - 適用する際のポイントはオラクルの設計
 - 系列ラベリング
 - Shift-reduce 依存構造解析
 - 大域学習と比較したときのメリット
 - 特徴量設計の自由度
 - 決定的解析(高速)でも高精度

既存手法: 正しい行動履歴を仮定した 決定的解析器の学習



既存手法の問題点

方策 π : 状態 $s \in S$ から行動 $a \in A$ への写像 $S \rightarrow A$

- 既存手法の目的関数⊖:
オラクルが観測する状態遷移下で損失関数の最小化

$$\arg \min_{\pi \in \Pi} E_{P_{\pi^*}(s)} [\text{loss}(\pi(s_i), \pi^*(s_i))] \quad \pi^*: \text{オラクル}$$

- オラクルと方策の状態分布の違い → 誤り伝播

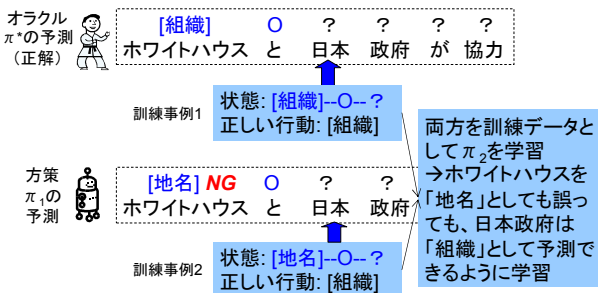
$$P_{\text{oracle}}(s) \neq P_{\text{policy}}(s)$$

- 模倣学習の目的関数⊖:
方策が観測する状態遷移下で損失関数を最小化

$$\arg \min_{\pi \in \Pi} E_{P_{\pi}(s)} [\text{loss}(\pi(s_i), \pi^*(s_i))]$$

模倣学習アルゴリズム DAGGER (Ross+ 2011)

過去の予測が間違える可能性を考慮した学習



DAGGER;

Dataset Aggregation (Ross+ 2011)

1. 現在の方策 π_k を実行し、観測された状態(特徴ベクトル)と正しい行動(正解ラベル)の組を収集(訓練集合 D_k)
 - 正しい行動を示すオラクル π^* の存在が前提
2. 過去に収集したすべての $D_1 \dots D_k$ を使って方策 π_{k+1} を学習 (1へ)

Algorithm 1 DAGGER
 初期化: $D \leftarrow \emptyset, \pi_1 \leftarrow \pi^*$
for $k = 1, 2, \dots, K$ **do**
 π_k を実行し $D_k = \{(s_t, \pi^*(s_t))\}$ を収集
 データを集約: $D \leftarrow D \cup D_k$
 D を用いて π_k を学習
end for
 検証用データで性能の良い π_k を選択

- 誤差上限は行動数 T に
対して線形 $O(T)$
- D_1 だけで学習 $\rightarrow O(T^2)$

固有表現抽出での検証結果 (CoNLL2002テスト性能)

- 長い予測履歴 (1,2,3,4-gram) を特徴にしても過学習しにくい
- F値: 模倣学習なし69前後 \rightarrow あり77以上 (CRFと同等)

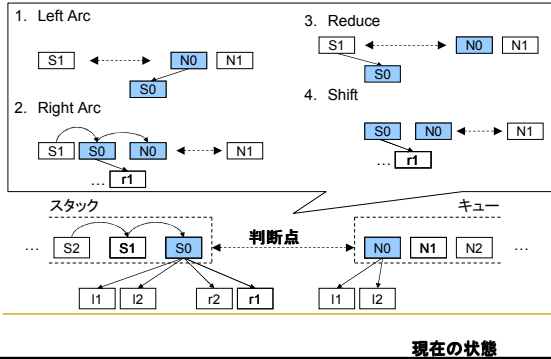
方策	DAGGER	精度	再現率	F値
ロジスティック回帰	なし	69.29	69.06	69.17
ロジスティック回帰	あり	77.49	77.01	77.25
SVM	なし	68.05	69.79	68.91
SVM	あり	77.93	77.68	77.80
CRF (一次のマルコフモデル)		78.04	77.09	77.56

オラクル π^* の設計

模倣学習を適用するポイント

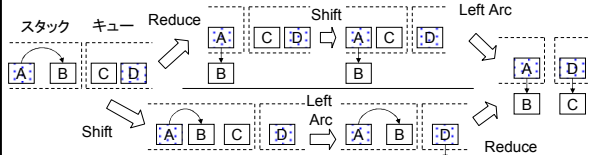
- オラクル π^* :
ある状態 s に対する最適行動を返すルール
 - 系列ラベリング問題でのオラクル(自明)
 - 状態にかかわらず予測点の正解ラベルを返す $\pi^*(s_i) = y_i$
 - 一般の構造予測問題でのオラクルの設計課題
 - 状態によっては最適行動が複数存在
 - 予測誤りによって回避不可能な誤りが発生
- \Rightarrow Shift-reduce法による決定的な依存構造解析

Shift-reduce 依存構造解析(arc-eager法, Nivre 2003)

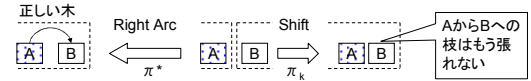


Shift-reduce 依存構造解析におけるオラクル設計の考慮点

- 別々の行動列が同じ木を構築可能

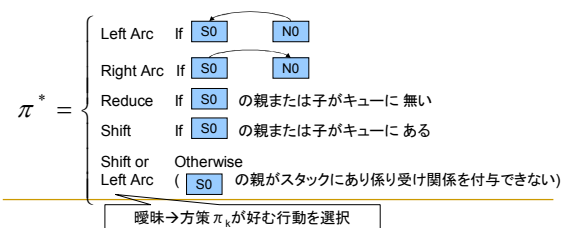


- 回復不可能な過去の誤り



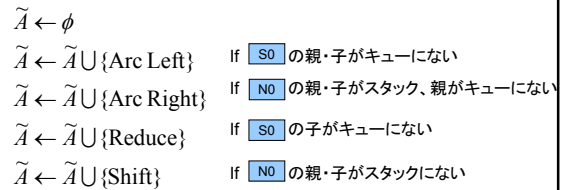
Shift-Reduce 依存構造解析のためのオラクル(曖昧性小)

- 予測誤りがない限り正規オラクルと同じ行動(曖昧性なし)
- 回避不可能な誤りがある場合のみ、将来の誤りが増えない行動を選択(曖昧性あり)



Shift-Reduce 依存構造解析のためのオラクル(曖昧性大) Goldberg&Nivre 2012

- 予測誤りがない場合も、将来の誤りが増えない行動の中から方策 π_k が好む行動を選択: $\pi^* = \arg \max_{a \in \tilde{A}} \pi_k(a|s)$



英語依存構造解析 Penn Treebank (Yamada+2003 head rule) テスト性能

- 方策=SVM、特徴量=Zhang&Nivre2011
- 評価指標
 - 単語単位: UAS:ラベルなし依存関係正答率, LAS:ラベルあり依存関係正答率
 - 文単位: UEM: ラベルなし完全一致

決定的解析での性能

オラクル曖昧性	DAGGER	ビーム幅	UAS	LAS	UEM
なし	なし	1	88.9	87.6	34.2
小さい	あり	1	90.0	88.7	34.7
大きい	あり	1	90.2	89.0	35.9
構造化パーセプトロン		1	88.6	87.1	34.0

模倣学習による決定的解析のメリット 大域学習(Global Training)との比較

- 動的計画法: 解析候補から最適解を選択できる一方、特徴量を小範囲に限定
→ 広範囲の予測履歴を特徴量として利用可能
- ビーム探索・リランキング: 広範囲の特徴量を使用する一方、解析候補から近似解を選択
→ ビーム探索無し(速い)でも高精度

オラクル曖昧性	DAGGER	ビーム幅	UAS	速度(文/秒)
なし	なし	1	88.9	1408
小	あり	1	90.0	1310
小	あり	2	90.5	637

ビーム幅2で解析速度が半減

まとめと今後の課題

- 高精度な決定的解析器が実現可能
- 今後の課題はオラクルの設計技法
 - 曖昧性の高いオラクルは目的関数の非凸性を高める
- 時間の都合で割愛した内容(予備スライド)
 - ビーム探索併用時のDAGGERの性能
 - 特徴量追加→決定的依存構造解析でUAS=90.7
 - その他の模倣学習アルゴリズム - オンラインアルゴリズム(Goldberg&Nivre 2012)やSEARN (Daume 2009)との比較

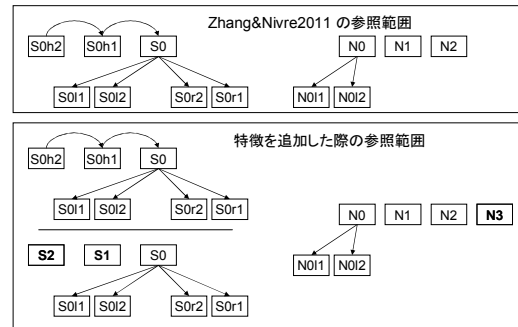
Backup slides

英語依存構造解析 テスト性能 ビーム探索と併用時

- ビーム幅=2での構造化パーセプトロンとの比較
 - 単語単位では模倣学習の性能が高いが、文単位では構造化パーセプトロンが勝る→模倣学習の目的関数は構造全体の予測を目指していないため

オラクル曖昧性	DAGGER	ビーム幅	UAS	LAS	UEM
小さい	あり	1	90.0	88.7	34.7
大きい	あり	1	90.2	89.0	35.9
小さい	あり	2	90.5	89.3	37.2
大きい	あり	2	90.6	89.5	37.3
構造化パーセプトロン		2	90.3	88.9	40.0
構造化パーセプトロン		64	92.8	91.7	48.0

Zhang&Nivre2011の特徴量の拡張



Zhang&Nivre2011の特徴量拡張の効果

オラクル曖昧性	DAGGER	特徴追加	ビーム幅	UAS	LAS	UEM
なし	なし	なし	1	88.9	87.6	34.2
小さい	あり	なし	1	90.0	88.7	34.7
大きい	あり	なし	1	90.2	89.0	35.9
なし	なし	あり	1	89.7	88.0	37.3
小さい	あり	あり	1	90.7	89.4	38.3
大きい	あり	あり	1	90.7	89.5	38.9
構造化パーセプトロン		あり	1	88.6	87.1	34.0
構造化パーセプトロン		あり	1	89.7	88.1	36.1

- 特徴追加により解析速度は1310文/秒→830文/秒に低下 (特徴追加なしビーム幅=2は 637文/秒)

オンライン模倣学習アルゴリズム (Goldberg&Nivre 2012)

Algorithm 2 Perceptron algorithm (Goldberg & Nivre 2012).

```

Set  $w \leftarrow 0$ 
for  $k = 1, 2, \dots, K$  do
  for sentence  $x$  with gold tree in corpus do
     $s$  is the start state for  $x$ 
    while  $s$  is not terminal do
       $a \sim \delta \pi_w(s) + (1 - \delta) \pi^*(s)$ 
       $\hat{A} \leftarrow \pi^*(s)$ 
       $a_p \leftarrow \operatorname{argmax}_{a \in A} w_a^\top \phi(s)$ 
      if  $a_p \notin \hat{A}$  then
         $a^* \leftarrow \operatorname{argmax}_{a \in A} w_a^\top \phi(s)$ 
         $w_p \leftarrow w_{a_p} - \phi(s)$ 
         $w_{a^*} \leftarrow w_{a^*} + \phi(s)$ 
      end if
       $s \leftarrow f(s, a)$ 
    end while
  end for
end for
Return  $\hat{w}$ .
    
```

次の行動は方策またはオラクルから確率的に選択(Stochastic mixed oracle)

$$\left(\max_{a \in A} w_a^\top \phi(s) - \max_{a \in \hat{A}} w_a^\top \phi(s) \right) \geq 0$$

各状態遷移ごとにパラメータ更新
→学習が速い

オンライン模倣学習アルゴリズム(Goldberg&Nivre 2012)およびSEARN (Daume III 2009) との比較

オラクル曖昧性	DAGGER	ビーム幅	UAS	LAS	UEM
なし	なし	1	88.9	87.6	34.2
小さい	あり	1	90.0	88.7	34.7
大きい	あり	1	90.2	89.0	35.9
Goldberg&Nivre 2012		1	89.9	88.7	34.2
Goldberg&Nivre 2012 (*1)		1	90.3	89.0	35.4
SEARN (オラクル曖昧小)		1	89.8	88.5	33.5
SEARN (オラクル曖昧小)(*2)		1	90.0	88.8	34.7
SEARN (オラクル曖昧大)(*2)		1	90.0	88.7	35.4

- *1) 処理順をランダムに変更
- *2) パラメータwを平均化

DAGGER 方策の平均化

Algorithm 1 DAGGER with policy averaging

```

Set  $D \leftarrow \emptyset, w \leftarrow 0, \bar{w} \leftarrow 0$ 
for  $k = 1, 2, \dots, K$  do
  if  $k = 1$  then
    Set  $\pi_k = \pi^*$ 
  else
    Set  $\pi_k(s) = \operatorname{argmax}_\pi w_{k,\pi}^\top \phi(s)$ 
  end if
  Collect data  $D_k = \{(\phi(s_{\pi_k}), \pi^*(s_{\pi_k}))\}$  by
  running  $\pi_k(s)$  on training set.
  Aggregate datasets:  $D \leftarrow D \cup D_k$ 
  Obtain  $w_{k+1}$  by optimizing  $w$  and  $\lambda$  for the
  empirical average of Eq. (1) over  $D$ 
  if  $k > 2$  then
    Update  $w_{k+1} \leftarrow \frac{k-1}{k} w_k + \frac{1}{k} w_{k+1}$ 
    else
    Update  $w_{k+1} \leftarrow w_{k+1}$ 
  end if
end for
Return best  $\bar{w}_k$  on validation.
    
```

$$\bar{w}_{k+1} = w_2 + w_3 + \dots + w_k$$

英語依存構造解析での模倣学習の効果 (数字は係り先番号、赤字は誤り)

単語	正解	DAGGER	SVM
It	2	2	2
would	24	0	0
be	2	2	2
my	5	5	5
inclination	3	3	3
to	7	7	7
advise	3	5	5
clients	7	7	7
not	11	7	7
to	11	11	11
sell	7	7	7
.	11	7	7
to	14	14	14
look	11	7	7
for	14	14	14
an	17	17	17
opportunity	15	15	15
to	19	19	19
buy	17	17	17
.	24	24	2
Mr.	23	23	23
Ball	24	24	24
told	0	0	2
the	26	26	26
brokers	24	24	24
.	24	24	2

(1) DAGGER, SVMとも係り先(親)を間違える("Would"を文のHeadと誤る)

"would" を文のHeadと誤った誤り(1)の影響でSVMは誤りを連発する
DAGGERは"told"を文のHeadとして再認識して誤り(1)の影響を受けない

英語依存構造解析 テスト性能

オラクル曖昧性	DAGGER	ビーム幅	UAS	LAS	UEM
なし	なし	1	88.9	87.6	34.2
小さい	あり	1	90.0	88.7	34.7
大きい	あり	1	90.2	89.0	35.9
小さい	あり	2	90.5	89.3	37.2
大きい	あり	2	90.6	89.5	37.3
構造化パーセプトロン		1	88.6	87.1	34.0
構造化パーセプトロン		2	90.3	88.9	40.0
構造化パーセプトロン		64	92.8	91.7	48.0

誤り伝播の例(固有表現抽出)

ラベル付き訓練事例 (地名, 組織, O(その他))

ラベル列
[地名]-O-[地名]
[地名]-[地名]
を重視して学習

[組織][組織] [組織][組織] O O	豊橋 技術 科学 大学 が 設立
[地名] O [地名] [地名][地名]	広島 と 名古屋 大学 駅
[地名] O [地名] O O	名古屋 と 京都 が 会場

決定的解析(テスト時)

NG NG
[地名] [地名] O ? ? ? ? ?
名古屋 大学 と 豊橋 技術 科学 大学 が 設立

誤り伝播の例(固有表現抽出)

ラベル列
[地名]-O-[地名]
[地名]-[地名]
を重視して学習

[組織][組織] [組織][組織] O O	豊橋 技術 科学 大学 が 設立
[地名] O [地名] [地名][地名]	広島 と 名古屋 大学 駅
[地名] O [地名] O O	名古屋 と 京都 が 会場

誤り伝播

NG NG NG NG
[地名] [地名] O [地名] [地名] [地名] [地名] O O
名古屋 大学 と 豊橋 技術 科学 大学 が 設立

Non i.i.d. (independent and identically distributed) data

- Wrong objective ☹️: minimizing a surrogate loss with respect to the oracle's state transition .

$$\arg \min_{\pi \in \Pi} \sum_t E_{s \sim d_{\pi^*}} [\text{loss}(\pi(s_t), \pi^*(s_t))] \quad \begin{array}{l} \pi^* \text{ : oracle} \\ d_{\pi^*}(s) = P_{\text{oracle}}(s) \end{array}$$

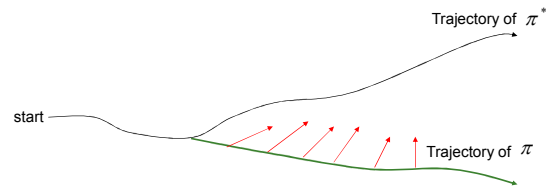
- Difference between state distribution of the oracle and the agent → error propagation and label bias

$$P_{\text{oracle}}(s) \neq P_{\text{agent}}(s)$$

- True objective ☺️: minimizing a loss with respect to the agent's state transition.

$$\arg \min_{\pi \in \Pi} J(\pi) \equiv \sum_t E_{s \sim d_{\pi}} [\text{loss}(\pi(s_t), \pi^*(s_t))]$$

Performance guarantee for DAGGER



Theorem 3. (Ross et al., 2011) For Dagger, if $Q_{T-t+1}^{\pi^*}(s, \pi) - Q_{T-t+1}^{\pi^*}(s, \pi^*) \leq u$ and N is $O(uT)$, there exists a policy $\pi \in \pi_{1:N}$ s.t. $J(\pi) \leq J(\pi^*) + uT\epsilon_N + O(1)$.

$$Q_i^{\pi^*}(s, \pi) = \text{loss}(\pi(s), \pi^*(s)) + Q_{i+1}^{\pi^*}(t(\pi(s)), \pi^*)$$

$$\epsilon_N = \arg \min_{\pi \in \Pi} \sum_{t=1}^T E_{s \sim d_{\pi_t}} [\text{loss}(\pi(s), \pi^*(s))]$$