

# 新語を含む単語分割のためのストリーム型能動学習における獲得関数の強化学習

脇一晟†

堅田俊†

武田龍†

駒谷和範†

†大阪大学 産業科学研究所

## 1. はじめに

対話システムにおいて新語に遭遇した際に、対話中にユーザーに確認を行うことで、システム自身で語彙を更新できると望ましい。それにより、対話後に管理者が語彙登録をする手間を省ける。ユーザーへの確認回数は、ユーザーに与える印象を考慮すると、最小限に留めたい。

このような問題設定はストリーム型能動学習 [1] として捉えることが出来る。ストリーム型能動学習では、ラベル付けされていない学習データが一つずつ順番に提示される。このストリームとして与えられるデータに対し、正解ラベルを要求してモデルの学習に用いるか否かを判断する。それによりアノテーションにかかる費用を抑えつつより良いモデルの学習を行う。

本研究では、新語を含む文の単語分割モデルの能動学習を題材として、その獲得関数を強化学習により得る事で、限られた正解データで性能の向上を目指す。ストリームとして与えられる文を学習に用いるべきか判断するために、学習途中の単語分割モデルを適用した際の確信度を獲得関数の入力に用いる。これは確信度の低いサンプルを選んで学習する、不確実性サンプリング [2] に基づいている。図1の1つ目の例では、ストリームから「イイテンキダネ」という文字列が提示されているが、分割結果の確信度が高いため、正解を要求しない。2つ目の例では、「マリトツツオダヨ」という文字列が提示されている。分割結果の確信度が低いため、正解データを要求し、単語分割モデルの学習に用いている。

## 2. 単語分割の能動学習

単語分割では、文字列  $x$  が与えられたときの、単語列  $w$  を推定する。これは、文字列  $x$  に対する単語境界変数列  $z$  を推定する問題に置き換えることが出来る。 $z$  はある文字列における単語の区切り位置を表すベクトルである。 $z$  は  $x$  と同じ次元数の 0 か 1 の要素を持ち、1 はその文字の後に単語区切りがあること、0 は単語区切りがないことを表す。単語境界変数列と文字列から、単語列が得られる。例えば、 $x = [ 'ト'; 'ウ'; 'ザ'; 'イ' ], z = [ 0, 1, 0, 1 ]$  が与えられたとき、 $w = [ 'トウ'; 'ザイ' ]$  となる。単語境界変数列を推定する問題は、予測結果を  $z'$  とすると、式 (1) のように記述できる。

$$z' = \underset{z}{\operatorname{argmax}} p(z|x) \quad (1)$$

本研究では、新語を含む文を追加学習する形で単語分割モデルのストリーム型能動学習を実施する。単語分割モデルは確率  $p(z|x)$  を学習し、これを用いて推定を行う。事前学習として一般的な語彙が含まれるコーパスでモデルを訓練しておく。事前学習したモデルでは、新語を含む文字列について、正しく推定することが出来ない。新語を含むデータを取捨選択して能動学習することで、新語を含む文を単語分割できるように訓練する。

Reinforcement Learning of Acquisition Function in Stream-based Active Learning for Segmenting New Words: Issei Waki, Shun Katada, Ryu Takeda, and Kazunori Komatani (Osaka Univ.)

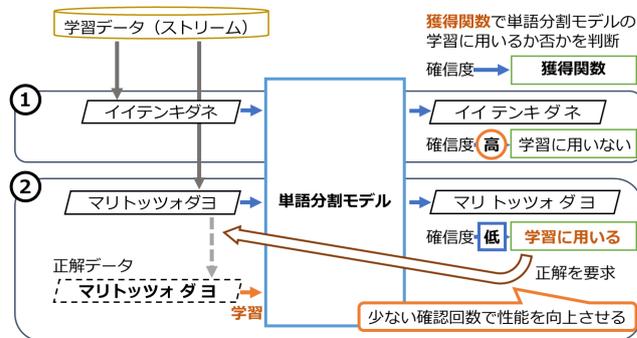


図1: ストリーム型能動学習による単語分割モデルの学習データ獲得

## 3. 獲得関数の強化学習

ストリーム型能動学習のプロセスをマルコフ決定過程に対応付けることで、獲得関数を強化学習により得る。これは Fang らの論文 [3] で固有表現抽出タスクにおいて提案された手法である。著者らはある言語で学習した獲得関数を別の言語における能動学習に用いている。本研究は単語分割タスクを扱う点で異なる。また学習した獲得関数は同一の言語に適用する。著作らの研究では文字列のベクトル表現が獲得関数の入力として有効だと述べられていた。しかし本研究のタスクには新語が含まれるため、文字列のベクトル表現を扱うことはできない。確信度と周辺確率のみを獲得関数の入力として扱う。

### 3.1 強化学習環境

状態はストリームから与えられた文字列  $x$  と現在の言語モデル  $\phi$  に依存する。状態は二種類あり、モデル  $\phi$  で文字列  $x$  の単語分割を行う際の結果を用いて導出する。一つ目は確信度の自然対数  $\log \max_z p_\phi(z|x)$  である。二つ目は文中に含まれる各文字の後に分割位置がある周辺確率  $p_\phi(z_i = 1|x)$  である。最大 50 文字分の数値を結合したベクトルとして扱う。次元数は 50 で固定し、文字数が足りない部分は 1 とした。

行動は獲得関数の出力であり、True か False のいずれかである。True はストリームから与えられた文を学習に用いることを表し、False は学習に用いないことを表す。

状態遷移はランダムなストリームから与えられる文と現在のモデルに依存する。

報酬は検証データによって求められる現在のモデルの精度の増減を用いる。精度の算出では完全に単語分割できた文のみを正解とする。

学習の終了条件は二種類ある。一つは規定個数 (Budget) のデータの学習が終了したとき、二つ目は学習データのストリームが空になったときである。

### 3.2 強化学習手法

言語モデルの追加学習を行う一連の流れを 1 エピソードとし、複数エピソード実施する。エピソードの初めに

言語モデルは事前学習したモデルにリセットされ、ストリームに用いる学習データもシャッフルされた後に、最初から読み込まれる。

戦略を学習する手法として Q-learning を用いる。これは状態に対する行動の価値、Q 値を予測・更新し、それに基づいて行動決定する手法である。Q 値を推定する Q 関数をニューラルネットワークにより回帰する。ストリームから提示される 1 つのデータ毎の処理を 1 エポックとし、エポック毎に Q 関数の更新をする。エポック毎にリプレイバッファに経験を格納し、またリプレイバッファを用いて Q 関数をミニバッチ学習する。

獲得関数は学習した Q 関数により出力が決定される。Q 関数は状態を受け取り二次元のベクトルを出力する。出力は二種の行動、True, False に対する Q 値である。獲得関数の出力として、Q 値が大きい方の行動を選択する。

#### 4. 評価実験

ベイジ階層言語モデルの一種である NPYLM (nested PitmanYor language modeling) [4] を用い、カナ文字の単語分割を行う。事前学習では「日本語の話し言葉コーパス」を用い教師なし学習を行った。能動学習ではコーパスに含まれない新語を含む対話文によるデータセットを用いた。データセットは合計 921 文であり、3.5:3.5:2:1 で無作為に分割し、学習データ、評価用学習データ、テストデータ、検証データとした。獲得関数の訓練では、学習データと検証データを用いた。評価実験では評価用学習データをストリームとして用いて、テストデータに対する分割精度の推移をベースラインと比較した。

##### 4.1 強化学習による獲得関数の訓練

強化学習により二種類の獲得関数を訓練した。一つ目は状態として確信度のみを用いるもの(確信度のみ)。二つ目は状態として確信度と周辺確率の両方を用いるもの(確信度と周辺確率)。

学習に使えるデータ数、Budget を 50 として、200 エピソードの学習を行った。割引率  $\gamma$  は 0.99、学習率は  $10^{-6}$  とした。epsilon-greedy 法を用いており、確率  $\epsilon$  で獲得関数に従い、そうでない場合はランダムな行動選択を取る。初期  $\epsilon$  は 0.95 で、100 エポックかけて 1 に近づけた。

Q 関数の学習には単層の線形ニューラルネットワークを用いた。状態ベクトルを入力すると、各行動の Q 値が出力として得られる。入力次元数は状態と同じで、出力次元数は 2、隠れ層は 64 である。リプレイバッファ容量は 100 で、ミニバッチのサイズは 32 とした。

##### 4.2 実験条件

強化学習により訓練した獲得関数で budget 数 50 の学習を行い、ベースライン二種類と比較した。一つはデータを学習に用いるか否かをランダムに決定するものであり(ランダムベースライン)、確率 0.2 で学習に用いる。もう一つは確信度が閾値より大きいか低いかでデータの要不要を決定するものである(確信度オラクル)。閾値は  $-0.51$  で、評価実験と同じ条件で探索を行い、最も精度が高いものを選んだ。オラクルに近い設定といえる。

##### 4.3 実験結果と考察

各獲得関数の評価実験におけるテストデータに対する精度を表 1 に示す。強化学習による手法については訓練時の精度も示している。また評価用学習データをストリームとして各々の獲得関数で得られた学習データに対する分割精度の推移を図 2 に示す。

表 1: 各獲得関数による学習終了時の精度 (学習サンプル数最大 50)

	強化学習時	評価実験時
確信度のみ	0.913	0.929
確信度と周辺確率	0.913	0.902
ランダムベースライン	-	0.896
確信度オラクル	-	0.934

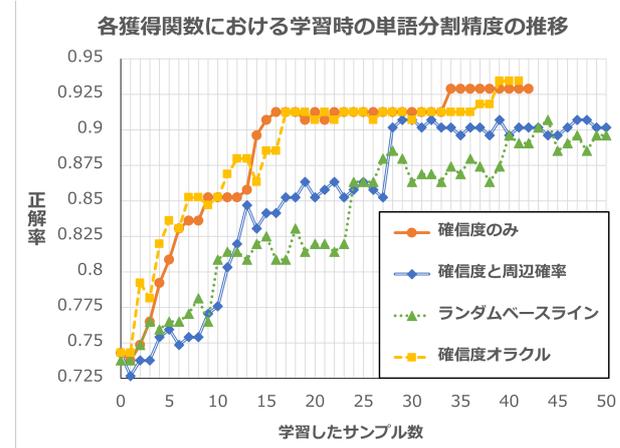


図 2: 評価用学習データを用いて学習した際の単語分割精度の推移

強化学習による手法は訓練時はどちらも同じ精度であったのに対し、評価実験時では確信度のみの方が高い精度となった。確信度と周辺確率設定における獲得関数の入力次元数が大きいことによる、強化学習時のストリームに対する過学習が原因とみられる。

確信度のみ設定はランダムベースラインより高い精度を示し、オラクルに近い閾値である確信度オラクルと同等の分割精度を示した。上位二つの獲得関数ではストリームを使い果たしたため、Budget に満たない個数のデータしか学習出来ていないが、高い分割精度を出している。学習データ量がより豊富なストリームの場合、更なる性能向上が見込める。

#### 5. おわりに

本稿ではカナ文字列単語分割タスクにストリーム型能動学習と獲得関数の強化学習を適用した。確信度を入力とした獲得関数を、強化学習を通じてある程度最適化出来ることを示した。今後は状態として用いる別の効果的な特徴量を模索し、少ないサンプル数でより性能の高い学習が行える獲得関数の構築を目指す。また、本研究では限定的な新語に対する学習しか出来なかったため、より大規模なデータセットを用いた実験も行いたい。

#### 参考文献

- [1] 日野 英逸: 能動学習: 問題設定と最近の話題. 日本統計学会誌, Vol. 50, No. 2, pp.317-342 (2021)
- [2] D.D.Lewis, and W.A.Gale: A sequential algorithm for training text classifiers. In Proc. ACM-SIGIR, pp.3-12 (1994)
- [3] M.Fang, Y.Li, and T.Cohn.: Learning how to Active Learn: A Deep Reinforcement Learning Approach. In Proc. EMNLP, pp.595-605 (2017)
- [4] D.Mochihashi, T.Yamada, and N.Ueda: Bayesian unsupervised word segmentation with nested PitmanYor language modeling. In Proc. ACL-IJCNLP, pp.100-108 (2009)