

RNNによる遷移確率計算を用いた隠れマルコフモデル

平岡 達也^{1,a)} 高瀬 翔^{1,b)} 内海 慶^{2,c)} 櫛 惇志^{2,d)} 岡崎 直観^{1,e)}

概要: 本研究では隠れ状態間の長距離依存を RNN で考慮し、遷移確率の計算を行う隠れマルコフモデルを提案する。従来のニューラルネットワークを用いた隠れマルコフモデルでは、隠れ状態の遷移確率を直前の時刻のみに依存して計算する。自然言語処理における品詞などの隠れ状態には長距離依存があると考えられるが、既存の手法ではこれを考慮できていない。そこで、RNN を用いて隠れ状態の系列をエンコードすることにより、隠れ状態の遷移をよりの確に捉えることができると考えられる。一方で、可能な全ての隠れ状態の系列について RNN を計算することは現実的ではないため、効率的な計算のための工夫が必要である。本研究では遷移確率を用いて RNN の隠れベクトルに重み付けを行うことで、隠れ状態の長距離依存を考慮した遷移確率計算が可能な隠れマルコフモデルを提案する。提案手法を教師なし品詞タグ付けによって実験・評価し、その結果と考察を報告する。

1. はじめに

自然言語などの系列データを用いた問題を解く上で、品詞やトピック遷移などの隠れた構造を的確に捉え活用することが重要である。こうした隠れ状態はアノテーションにより明示的にモデルへ与えられるだけではなく [1], [3], [9], モデル自身が生データから隠れ状態を推定する教師なし学習としても研究されてきた [12], [13], [15].

生データから隠れ状態を推定するために、これまで隠れマルコフモデルが広く用いられてきた。近年では、ニューラルネットワークを用いた隠れマルコフモデル (NHMM) が提案され、教師なし品詞タグ付けや要約文生成のテンプレート学習などにおいて性能の向上が報告されている [5], [14], [17]. NHMM では遷移確率や出力確率をニューラルネットワークで計算するため、単語列などの観測系列の情報を効果的に用いた学習が可能となる。例えば観測された単語列に含まれる文字情報を利用することで、出力確率の計算性能が向上すると考えられる。これまでの NHMM は単純マルコフ過程をベースとしており、遷移確率の計算にはそれまでの観測値や直前の隠れ状態を用いている。

自然言語処理における品詞などの隠れ状態には、隠れ状態同士の長距離依存があると考えられ、これを考慮する

ことでよりの確に隠れ状態の推定ができると期待される。そこで本研究では、リカレントニューラルネットワーク (RNN) に隠れ状態を入力することで隠れ状態間の長距離依存を表現し、各時刻の遷移確率を計算する新たなモデルを提案する。

実際の隠れマルコフモデルをベースとした計算において、隠れ状態間の長距離依存を明に扱うためには、全ての可能な系列を展開し、各系列ごとに RNN の計算を行う必要がある。単純には可能な全ての系列に対応する RNN の計算結果を保持する必要があるが、一文の長さや隠れ状態の数を考えるとその組み合わせは膨大であり、現実的な手法とは言えない。

現実的には、より確からしい隠れ状態の遷移の情報に重きを置いた RNN の隠れベクトルの計算を行うことが望ましい。そこで提案手法では図 1 のように、単純マルコフ過程による隠れ状態の遷移確率を用いて RNN の隠れベクトルに重み付けを行う。これにより、RNN の隠れベクトルに隠れ状態間の長距離依存関係が反映されると仮定する。

このように拡張した RNN の隠れベクトルを用いることで、隠れ状態の系列を用いた遷移確率の計算を試みる。提案手法ではこれに加え、各時刻までの観測系列の双方向文脈を用いて遷移確率を計算する。

これまでの NHMM とは異なり、提案手法は前向き文脈と後ろ向き文脈の双方を利用している点、隠れ状態の系列を RNN を用いて直接的に考慮している点で新しく、各時刻の遷移確率の計算が効果的に行われると期待される。提案手法の有効性を NHMM の評価として広く用いられている教師なし品詞タグ付けで検証し、その結果を報告する。

¹ 東京工業大学

² デンソーアイティラボラトリ

a) tatsuya.hiraoka@nlp.c.titech.ac.jp

b) sho.takase@nlp.c.titech.ac.jp

c) kuchiumi@d-itlab.co.jp

d) akeyaki@d-itlab.co.jp

e) okazaki@c.titech.ac.jp

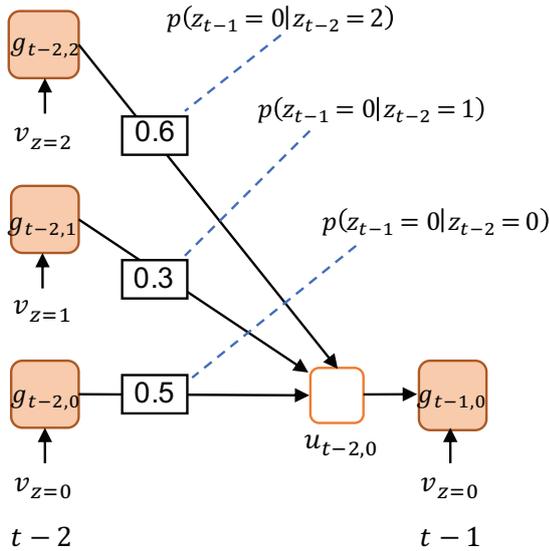


図 1 遷移確率による RNN の隠れベクトルの重み付け計算の概要. 時刻 $t-2$ の RNN に各隠れ状態に対応する分散表現 v_z が入力され, その計算結果を時刻 $t-1$ への遷移確率によって重み付けする. 次の時刻の RNN は, 前時刻の隠れベクトルの重み付け和を利用して計算を行う.

2. ニューラル隠れマルコフモデル

提案手法では既存の NHMM をベースとした拡張を行う. 本節では Tran らにより提案されたニューラルネットワークを用いた隠れマルコフモデル [14] について説明する.

通常の隠れマルコフモデルでは, 観測された系列 $\mathbf{x} = x_1 \dots x_t \dots x_T$ を生成する隠れ状態の系列 $\mathbf{z} = z_1 \dots z_t \dots z_T$ を求める. 隠れ状態の系列と観測値の系列との同時確率は以下の式で表される.

$$p(\mathbf{x}, \mathbf{z}) = \left(\prod_{t=2}^T p(z_t | z_{t-1}) \right) \left(\prod_{t=1}^T p(x_t | z_t) \right) \quad (1)$$

Tran らによる NHMM では, ここで用いられる各確率の計算にニューラルネットワークを用いる. 例えば時刻 t におけるある隠れ状態 z_t から観測値 x_t が出力される確率 $p(x_t | z_t)$ は, 隠れ状態 z_t に対応する出力確率を計算するための分散表現 $\mathbf{v}_{z_t}^{\text{emis}}$ を用いて次のように計算される.

$$p(x_t | z_t) = \text{softmax}(\mathbf{W}^{\text{emis}} \mathbf{v}_{z_t}^{\text{emis}} + \mathbf{b}^{\text{emis}})_{x_t} \quad (2)$$

ここで \mathbf{W}^{emis} , \mathbf{b}^{emis} はパラメーター, $(\cdot)_{x_t}$ はベクトルの x_t に対応する要素を抜き出す操作である.

また, 各時刻での遷移確率 $p(z_t | z_{t-1})$ を表す遷移行列は, あるクエリベクトル \mathbf{q}_{t-1} を用いて次のように求めたスコア行列 $\mathbf{M}_{t-1} = (\mathbf{m}_{t-1,1}, \dots, \mathbf{m}_{t-1,j}, \dots, \mathbf{m}_{t-1,|z|})$ を列ごとに softmax を取ることで得る.

$$\mathbf{M}_{t-1} = \mathbf{W} \mathbf{q}_{t-1} + \mathbf{b} \quad (3)$$

$$p(z_t = i | z_{t-1} = j) = \text{softmax}(\mathbf{m}_{t-1,j})_i \quad (4)$$

Tran らはこのクエリベクトル \mathbf{q}_{t-1} の計算に, 観測された系列を Long-Short Term Memory (LSTM) [6] でエンコードしたベクトルを利用することで, 性能の向上を報告している [14].

$$\mathbf{q}_{t-1}^{\text{obs}} = \vec{\mathbf{h}}_{\leq t-1} \quad (5)$$

ここで, $\vec{\mathbf{h}}_{\leq t-1}$ は時刻 $t-1$ までの観測値をエンコードした LSTM の出力である. このクエリベクトル $\mathbf{q}_{t-1}^{\text{obs}}$ を用いて, 直前の隠れ状態が $z_{t-1} = j$ であるときのスコアベクトル $\mathbf{m}_{t-1,j}^{\text{obs}}$ を計算する.

$$\mathbf{m}_{t-1,j}^{\text{obs}} = \mathbf{W}_j^{\text{obs}} \mathbf{q}_{t-1}^{\text{obs}} + \mathbf{b}_j^{\text{obs}} \quad (6)$$

さらに直前の隠れ状態 $z_{t-1} = j$ から $z_t = i$ へと移行する確率は, 以下のように計算される.

$$p(z_t = i | z_{t-1} = j) = \text{softmax}(\mathbf{m}_{t-1,j}^{\text{obs}})_i \quad (7)$$

このようにして計算された出力確率 $p(x_t | z_t)$ および遷移確率 $p(z_t | z_{t-1})$ を利用して前向き・後ろ向きメッセージ α および β を以下のように計算する.

$$\alpha_{t,z_t} = p(x_t | z_t) \sum_{z_{t-1}} p(z_t | z_{t-1}) \alpha_{t-1,z_{t-1}} \quad (8)$$

$$\beta_{t,z_t} = \sum_{z_{t+1}} p(z_{t+1} | z_t) p(x_{t+1} | z_{t+1}) \beta_{t+1,z_{t+1}} \quad (9)$$

ここで前向きメッセージの初期状態は一様分布, 後ろ向きメッセージの初期状態は 1 である. 計算された α と β を用いて, 事後確率 $p(z_t = i | \mathbf{x})$ が次のように計算される.

$$p(z_t = i | \mathbf{x}) \propto \alpha_{t,z_t=i} \beta_{t,z_t=i} \quad (10)$$

これにより, 観測系列に対して確率が最大となる隠れ状態を推論することができる.

3. 提案手法

提案手法による遷移確率の計算の概要を図 2 に示した. 本論文では, 既存の NHMM における遷移確率の計算に二つの拡張を施す. 一つ目に, 前向き方向の観測系列に加えて後ろ向き方向の観測系列を考慮した遷移確率の計算を提案する (図 2 右). 二つ目に, RNN を用いて隠れ状態の系列をエンコードし, 隠れ状態間の長距離依存を考慮した遷移確率の計算を提案する (図 2 左). それぞれの具体的な計算方法について, 順に説明する.

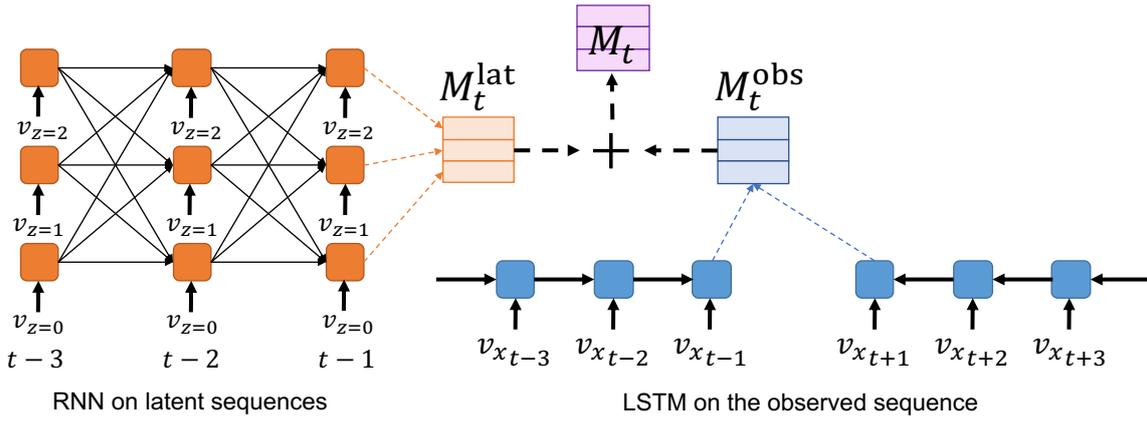


図2 提案手法によるスコア行列計算の概要. 提案手法は双方向の LSTM による観測値の分散表現 v_x の系列をエンコードした結果 (図右) と RNN で隠れ状態の分散表現 v_z の系列をエンコードした結果 (図左) を用いて, 遷移スコア行列を計算する.

3.1 双方向文脈を利用した遷移確率計算

NHMM による隠れ状態の推論を行うとき, 全ての観測系列の情報を利用して隠れ状態の系列を効果的に計算できると期待される. そこで提案手法では, Tran らにより導入された前向き文脈をエンコードする LSTM に加え, 後ろ向き文脈を考慮する LSTM を導入することで, より多くの情報を用いて遷移確率の計算を行う.

提案手法では LSTM を用いたクエリベクトル q_t^{obs} の計算を以下のように拡張する.

$$q_{t-1}^{\text{obs}} = \vec{h}_{\leq t-1} \oplus \overleftarrow{h}_{\geq t+1} \quad (11)$$

ここで $\overleftarrow{h}_{\geq t+1}$ は後ろ向き LSTM によって時刻 T から $t+1$ までの観測値をエンコードしたベクトルを表す. また \oplus はベクトル結合を表し, 観測系列の前向き・後ろ向き文脈をエンコードしたベクトルを結合することでクエリベクトル q_{t-1}^{obs} を計算する. さらにこれを用いて, (6) と同様にスコア行列 $M_{t-1}^{\text{obs}} = (m_{t-1,1}^{\text{obs}}, \dots, m_{t-1,j}^{\text{obs}}, \dots, m_{t-1,|z|}^{\text{obs}})$ を計算する. この計算の概要を図2の右に示した.

3.2 RNN による隠れ状態の系列を考慮した遷移確率計算

ある時刻 t における遷移確率 $p(z_t|z_{t-1})$ を計算する際に, 隠れ状態の系列を観測値と同様にして RNN でエンコードすることで, 隠れ状態同士の長距離依存を考慮した遷移確率の計算が可能となる. 例えば時刻 t に i 番目の隠れ状態に対応する繊維確率の計算のための分散表現 $v_{z_t=i}$ が入力された RNN の隠れベクトル g_t は Elman ネットワークであれば以下のように計算される [4].

$$g_t = \sigma(\mathbf{W}^{\text{RNN}}(g_{t-1} \oplus v_{z_t=i}) + \mathbf{b}^{\text{RNN}}) \quad (12)$$

ここで σ は活性化関数であり, 本論文で述べる RNN では \tanh 関数を用いる. このように, ある一つの隠れ状態の系列に対しては, 素朴な RNN による系列のエンコード

が可能である. しかし実際の隠れマルコフモデルの計算において, 隠れ状態間の長距離依存を考慮するためには全ての可能な隠れ状態の系列について RNN の計算を行わなければならない.

単純には, 可能な全ての隠れ状態系列の組み合わせに対して式 (12) によるエンコードを行えばよい. しかし自然言語などの系列データの長さや, 隠れ状態の数を考えるとこの方法は現実的ではない.

そこで本研究では, 遷移確率を用いて RNN の隠れベクトルを重み付けすることで, ある時刻までの隠れ状態系列の組み合わせを隠れベクトルに保存する手法を提案する. ある時刻 $t-1$ から t への各隠れ状態への遷移確率を計算するとき, 以下のように直前の時刻で計算された遷移確率 $p(z_{t-1}|z_{t-2})$ を用いて, 時刻 $t-2$ に隠れ状態 $z_{t-2} = k$ が入力された RNN の隠れベクトル $g_{t-2,k}$ に重み付けを行う.

$$u_{t-2,j} = \sum_k p(z_{t-1} = j | z_{t-2} = k) g_{t-2,k} \quad (13)$$

時刻 $t-2$ の隠れベクトルの重み付け和である $u_{t-2,j}$ は, 時刻 $t-1$ に入力される隠れ状態ごとに計算される. 次に, $u_{t-2,j}$ を時刻 $t-1$ に隠れ状態 $z_{t-1} = j$ が入力された RNN の計算に用いる.

$$g_{t-1,j} = \sigma(\mathbf{W}^{\text{RNN}}(u_{t-2,j} \oplus v_{z_{t-1}=j}) + \mathbf{b}^{\text{RNN}}) \quad (14)$$

これは式 (12) で示した RNN の計算を拡張したものであり, 前時刻からの入力として遷移確率で重みづけられた隠れ状態の和 $u_{t-2,j}$ を用いた機構である. 図1では, 3つの隠れ状態をもつ NHMM による計算を示しており, 時刻 $t-1$ の隠れ状態が $j=0$ となる RNN の計算の概要を示している.

最後に RNN の隠れ状態をクエリベクトルとして, 時刻

$t-1$ の隠れ状態が j であるときの遷移確率を計算するためのスコアベクトルを以下のように計算する。

$$\mathbf{m}_{t-1,j}^{\text{lat}} = \mathbf{W}^{\text{lat}} \mathbf{g}_{t-1,j} + \mathbf{b}^{\text{lat}} \quad (15)$$

計算されたスコアベクトルを束ねたスコア行列 $\mathbf{M}_{t-1}^{\text{lat}} = (\mathbf{m}_{t-1,1}^{\text{lat}}, \dots, \mathbf{m}_{t-1,j}^{\text{lat}}, \dots, \mathbf{m}_{t-1,|z|}^{\text{lat}})$ を、観測値のエンコードによって計算されたスコア行列 $\mathbf{M}_{t-1}^{\text{obs}}$ と足し合わせて提案手法のスコア行列 \mathbf{M}_{t-1} を得る。

$$\mathbf{M}_{t-1} = \mathbf{M}_{t-1}^{\text{obs}} + \mathbf{M}_{t-1}^{\text{lat}} \quad (16)$$

この \mathbf{M}_{t-1} に対して式 (4) と同様、列ごとに softmax 計算を行うことで、時刻 $t-1$ から t における遷移行列を求める。このスコア行列の計算概要を図 2 に示した。

3.3 学習

教師なし学習では、実際の隠れ状態の系列 \mathbf{z} が不明であるため、Tran ら [14] と同様 Baum-Welch アルゴリズム [16] をベースとした学習を行う。この学習では、以下のように事後確率を用いて損失に重み付けを行うことで、勾配を期待値に基づいて再分配する。

$$J(\theta) = \sum_t \sum_{z_t} p(z_t|\mathbf{x}) \frac{\partial \ln p(x_t|z_t, \theta)}{\partial \theta} + \sum_{z_{t-1}} p(z_t, z_{t-1}|\mathbf{x}) \frac{\partial \ln p(z_t|z_{t-1}, \theta)}{\partial \theta} \quad (17)$$

式中の $p(z_t|\mathbf{x})$ は前向き後ろ向きメッセージを利用して式 (10) から求められ、 $p(z_t, z_{t-1}|\mathbf{x})$ についても以下のように求めることができる。

$$p(z_t = i, z_{t-1} = j|\mathbf{x}) \propto \alpha_{t,i} p(z_{t+1} = j|z_t = i) \times \beta_{t+1,j} p(x_{t+1}|z_{t+1} = j) \quad (18)$$

以上の損失関数を用いて勾配を計算し、Adam[8] によるパラメータの更新を行う。

4. 実験

4.1 データセット

提案手法の有効性を評価するために、隠れマルコフモデルの検証に広く用いられている教師なし品詞タグ付けタスクを用いる。本論文では、教師なし品詞タグ付けタスクにおいて一般的に用いられる Penn Tree Bank[10] の英語 WSJ コーパスによる学習を行う。性能の評価には、コーパスに付与された 45 種類の品詞タグを用いる。また、品詞タグの種類数を 12 に削減した Universal Treebank Version 2.0*[11] を用いて、タグ数が少ない状況でのモデルの評価

*1 <https://github.com/ryanmcd/uni-dep-tb>

もあわせて行う。

先行研究 [14] と同様、全ての数字を 0 に置き換える前処理のみ行った。また、今回比較するベースライン [5], [13], [14] と実験設定を揃えるため、45 タグの実験設定では WSJ コーパス全体での学習・評価を行い、12 タグの実験設定では学習用コーパスで学習、テストデータでの評価を行なった。

4.2 実験設定

全ての実験において、NHMM の隠れ状態数には正解となるタグ種類数と同じ 45 および 12 を用いる。観測値のエンコードに用いる単語分散表現とタグ分散表現のサイズは 128、LSTM と提案手法による RNN の隠れベクトルのサイズは 256 とした。また、観測値のエンコードに用いる LSTM の計算にはドロップアウトを 0.5 で設定した。学習には全ての長さの文を用い、バッチサイズは 32 とした。

ニューラルネットワークを用いたベースライン手法として、遷移確率行列をパラメータとして持ち学習するモデル (“NHMM”)、観測系列を LSTM でエンコードして遷移確率行列を計算するモデル (“+ forward LSTM”) を用いる。またこれに加え、本論文で提案する観測系列の後ろ側文脈をエンコードして遷移確率計算を行う手法 (3.1 節) を “+ forward&backward LSTM” として示した。なお “+ forward LSTM*” の設定は、直接の先行研究である Tran らの手法 [14] から出力確率に関する Convolution 層を除いたモデルと同じである。本論文で提案する隠れ状態系列についての RNN(3.2 節) を用いた手法は “+ hidRNN” として示す。隠れ状態系列の計算と出力確率の計算では、異なるタグ分散表現を用いた。

ニューラルネットワークを用いた手法の学習性能は初期値に強く依存するため、観測系列や隠れ状態系列のエンコードを行う手法の学習では、最もシンプルなニューラルモデルである “NHMM” による学習結果を初期値とした。ここで初期化されるパラメータは出力確率を計算する線形層とタグ埋め込み表現である。提案手法では出力確率の計算と遷移確率の計算について、それぞれ異なるタグ埋め込み表現を用いているため、その両方を同じ値で初期化し、それぞれ学習を行った。

評価には Many-to-One(M-1) 指標と、One-to-One(1-1) 指標を用いた。M-1 は品詞が複数の各隠れ状態に対応づけられることを許し、1-1 ではこれを許さず品詞と隠れ状態を貪欲に対応づける指標である。

4.3 結果

4.3.1 WSJ-12 タグ

WSJ コーパスの 12 タグ設定における実験結果を表 4.3.1 に示した。12 タグの実験と比較するベースラインモデルとして、通常の HMM を Baum-Welch アルゴリズムにより学習した手法 (HMM) および、Stratos らの手法のうち人

モデル	M-1(%)	1-1(%)
HMM[13]	59.8	-
ANCHOR[13]	66.1	-
NHMM	63.7	50.85
+ forward LSTM*	68.59	54.54
+ hidRNN	68.82	54.69
+ forward&backward LSTM	67.15	55.7
+ hidRNN(full proposed model)	70.22	58.48

表 1 Universal Treebank Version 2.0 の英語 WSJ コーパスにおけるタグ数 12 の教師なし品詞タグ付けの実験結果. HMM, ANCHOR の値は Stratos ら [13] によって報告された値を引用. *で示したモデルは直接の先行研究となるモデル [14] のうち, LSTM のみを用いたものと同じ設定.

手で作成された素性を用いないモデル (ANCHOR) をそれぞれ文献 [13] から引用し掲載した.

結果より, 提案手法による隠れ状態系列のエンコードを行う (+ hidRNN) ことで, M-1, 1-1 指標の両方で性能の向上が確認された. また, 隠れ状態系列のエンコードに加えて後ろ向き文脈を考慮することで, さらなる性能の向上が得られた. 前向き文脈のみを用いていた既存手法に後ろ向き文脈を追加した場合に M-1 指標では性能の低下が見られるが, 1-1 指標では性能が向上しているため, モデルの性能に大きな影響を与えていないと考えられる.

4.3.2 WSJ-45 タグ

WSJ コーパスの 45 タグ設定における実験結果を表 4.3.2 に示した. 45 タグの実験で比較するベースラインモデルとしては, 直接の先行研究である Tran らの手法 [14] および, 本データセットにおける現時点での最高性能である He らの手法 [5] を引用した. 表上段に示した “NHMM[14]**” および “NHMM + LSTM*” と下段に示した “NHMM**” および “NHMM + forward LSTM*” はそれぞれ同じモデル構造であるが, パラメータ数や初期値などが先行研究で報告された結果と異なるため個別に示した.

結果より, WSJ-12 タグ設定における実験とは異なり, 45 タグ設定での実験では後ろ向き文脈を考慮することで既存手法, 提案手法ともに性能の大幅な改善が見られた. タグ数が増え, より複雑な隠れ状態の遷移が起きる状況において, 後ろ向き文脈の情報が遷移確率の計算に良い影響を与えたと考えられる.

本研究では HMM の隠れ状態の系列を RNN でエンコードすることで遷移確率の計算を改善する手法を提案している. そのため, 既存手法のうち遷移確率の計算を改善している “NHMM + LSTM*” の性能が直接の比較対象である. 実験結果より, 提案手法による後ろ向き文脈と隠れ状態の系列のエンコードを盛り込むことで, M-1 指標での性能の向上が確認された. 一方で 1-1 指標では既存手法の数値に及んでいない. 既存手法と同じモデル構造である “NHMM + forward LSTM*” が既存手法の数値を再現できていないため, ハイパーパラメータや学習の初期値による差異で

モデル	M-1(%)	1-1(%)
HMM[14]	62.5	41.4
NHMM[14]**	59.8	45.7
+ LSTM*	65.1	52.4
+ Conv	74.1	48.3
+ Conv&LSTM	79.1	60.7
He ら, 2018[5]	80.8	-
NHMM**	66.67	47.28
+ forward LSTM*	68.05	44.73
+ hidRNN	68.49	45.04
+ forward&backward LSTM	72.01	48.39
+ hidRNN(full proposed model)	73.4	49.9

表 2 Penn Tree Bank の英語 WSJ コーパスにおけるタグ数 45 の教師なし品詞タグ付けの実験結果. *および**はそれぞれ同じモデル設定であることを示す.

あると考えられる.

Tran らの手法における “+ Conv” と He らの手法はどちらも出力確率の計算を改善しており, 本論文の提案手法はその性能に及んでいない. しかしこれらの出力確率計算の改善は提案手法にも盛り込むことができるため, これによりさらなる性能の向上が見込まれる.

5. 関連研究

自然言語処理において隠れマルコフモデルは広く研究され, その多くが教師なし品詞タグ付けタスクによって評価されてきた [2], [7], [15]. 近年の隠れマルコフモデルを用いた教師なし品詞タグ付けとして, 各隠れ状態に固有の単語が一つ存在するという制約を用いた Stratos らによる研究が挙げられる [13].

Tran ら [14] は隠れマルコフモデルの出力確率と遷移確率の計算にニューラルネットワークを用いる手法を提案し, Baum-Welch アルゴリズムをベースとした勾配の再分配による学習方法を提案した. 本研究はこの手法を基礎とし, 拡張したモデルを提案している. また, He らによる研究 [5] では, 離散的な隠れマルコフモデルの出力確率に単語分散表現を対応づけることで, 大幅な性能の向上を報告している.

6. まとめ

本論文では, あらゆる隠れ状態の系列を直接考慮して遷移確率計算を行える NHMM について提案した. 提案手法における NHMM では, 隠れ状態系列のエンコードに RNN を用い, この RNN の内部状態を遷移確率で重み付けすることで, あらゆる隠れ状態の系列を考慮できるよう拡張を行った. 提案手法は英語 WSJ コーパスにおける教師なし品詞タグ付けタスクによって評価し, 既存手法のうち遷移確率計算の改善を行なったモデルと比較して性能の向上が確認された. 一方で, 既存手法のうち出力確率計算の改善を行なった手法の性能には及んでおらず, さらなる手法

の改善が必要である。本論文で報告したモデルでは遷移確率の計算のみ拡張を行っており、既存研究で用いられた出力確率計算の拡張機構を用いていない。これらを活用して提案手法を拡張することで、性能のさらなる向上が見込まれる。

手法の改善に加え、提案手法の評価方法についてもさらなる発展が必要である。今後は多言語での教師なし品詞タグ付けタスクにおける評価や、文生成タスクにおけるテンプレートの自動学習などの応用タスクなどに提案手法を適用することで、詳しく評価、分析を行っていく。

謝辞 本研究成果は、国立研究開発法人情報通信研究機構（NICT）の委託研究により得られたものです。

参考文献

- [1] Allan, J. and Raghavan, H.: Using part-of-speech patterns to reduce query ambiguity, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 307–314 (2002).
- [2] Blunsom, P. and Cohn, T.: A hierarchical Pitman-Yor process HMM for unsupervised part of speech induction, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, pp. 865–874 (2011).
- [3] Collobert, R. and Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning, *Proceedings of the 25th international conference on Machine learning*, ACM, pp. 160–167 (2008).
- [4] Elman, J. L.: Finding structure in time, *Cognitive science*, Vol. 14, No. 2, pp. 179–211 (1990).
- [5] He, J., Neubig, G. and Berg-Kirkpatrick, T.: Unsupervised Learning of Syntactic Structure with Invertible Neural Projections, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1292–1302 (2018).
- [6] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
- [7] Johnson, M.: Why doesn't EM find good HMM POS-taggers?, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 296–305 (2007).
- [8] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [9] Kiperwasser, E. and Goldberg, Y.: Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations, *Transactions of the Association for Computational Linguistics*, Vol. 4, No. 1, pp. 313–327 (2016).
- [10] Marcus, M., Santorini, B. and Marcinkiewicz, M. A.: Building a large annotated corpus of English: The Penn Treebank (1993).
- [11] McDonald, R., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Oscar, T. et al.: Universal dependency annotation for multilingual parsing, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 92–97 (2013).
- [12] Shao, X., Xu, C. and Kankanhalli, M. S.: Unsupervised classification of music genre using hidden markov model, *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, Vol. 3, IEEE, pp. 2023–2026 (2004).
- [13] Stratos, K., Collins, M. and Hsu, D.: Unsupervised Part-Of-Speech Tagging with Anchor Hidden Markov Models, *Transactions of the Association for Computational Linguistics*, Vol. 4, No. 1, pp. 245–257 (2016).
- [14] Tran, K. M., Bisk, Y., Vaswani, A., Marcu, D. and Knight, K.: Unsupervised Neural Hidden Markov Models, *Proceedings of the Workshop on Structured Prediction for NLP*, pp. 63–71 (2016).
- [15] Van Gael, J., Vlachos, A. and Ghahramani, Z.: The infinite HMM for unsupervised PoS tagging, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, Association for Computational Linguistics, pp. 678–687 (2009).
- [16] Welch, L. R.: Hidden Markov models and the Baum-Welch algorithm, *IEEE Information Theory Society Newsletter*, Vol. 53, No. 4, pp. 10–13 (2003).
- [17] Wiseman, S., Shieber, S. and Rush, A.: Learning Neural Templates for Text Generation, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3174–3187 (2018).