

H-002

階層隠れ CRF の提案

Hierarchical Hidden Conditional Random Fields

杉浦 玄朗†
Takaaki Sugiura林 朗†
Akira Hayashi後藤 直人†
Naoto Goto末松 伸朗†
Nobuo Suematsu岩田 一貴†
Kazunori Iwata

1 はじめに

近年、生成モデルである隠れマルコフモデル (HMM) に代わる新たな手法として、識別モデルである Conditional Random Fields (CRF) が提案され、その有効性が確認されている [1]. CRF と HMM の違いは大きく 2 つある. 1 つ目は、最尤推定において、生成モデルである HMM は入力と出力の同時確率を推定するが、識別モデルである CRF は入力を条件付きとした出力確率を直接推定する点である. 2 つ目は、HMM は各時刻の入力と他時刻の入力との独立を仮定するが、CRF はこの仮定が必要ない点である.

本研究では、階層 HMM (HHMM) [2] に対応する識別モデル、階層隠れ CRF (HHCRF) を提案する. HHCRF の推論アルゴリズムとして、上層のラベルのみからの学習アルゴリズム、下層の状態系列を周辺化し上層の最適系列を求める周辺化ビタールゴリズムの 2 つを考案した. 比較実験により、HHCRF の優位性を示す.

2 HHCRF

2.1 モデル

HHCRF は条件付き確率分布 $p(Q^1|O; \Lambda)$ を表現する無向グラフィカルモデル (図 1) である. ここで、 $O = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T\}$ を時系列観測データ、 $Q^1 = \{q_1^1, q_2^1, \dots, q_T^1\}$ を時系列観測データに対応する第 1 層のラベル系列、 $Q^2 = \{q_1^2, q_2^2, \dots, q_T^2\}$ を第 2 層の隠れ状態系列、 $F = \{f_1, f_2, \dots, f_T\}$ を指標変数系列 (q_t^2 が終了状態に遷移すれば $f_t = 1$, それ以外ならば $f_t = 0$ とする変数系列)、 $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ をモデルパラメータとする. なお、本モデルは何層にも対応できるモデルであるが、スペースの都合上、2 階層隠れ CRF として説明する.

$$p(Q^1|O; \Lambda) = \frac{1}{Z(O; \Lambda)} \sum_{F, Q^2} \exp \left(\sum_{k=1}^K \lambda_k \phi_k(Q^1, F, Q^2, O) \right) \quad (1)$$

ここで、 $Z(O; \Lambda)$ は確率和を 1 にするための正規化項である.

$$Z(O; \Lambda) = \sum_{Q^1} \sum_{F, Q^2} \exp \left(\sum_{k=1}^K \lambda_k \phi_k(Q^1, F, Q^2, O) \right)$$

ϕ_k は素性関数、以下素性と呼ばれ、時間において連続な変数組 (例えば $\mathbf{o}_{t-1}, \mathbf{o}_t$) に対する特徴関数として定義される. 一

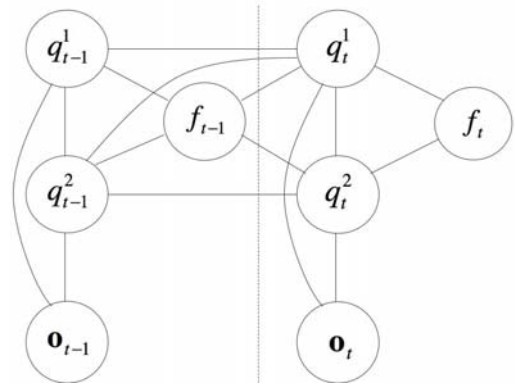


図 1 HHCRF の無向グラフ (時刻 $t-1, t$ に関する部分のみ)

般的に素性の選択は任意であるが、本研究ではガウス分布を出力する HHMM との性能を比較するため以下の素性を選択し、HHMM のモデル構造と等価なものにする. なお、 λ_k は各素性の重みである.

$$\begin{aligned} \phi_i^{(Prior)}(Q^1, F, Q^2, O) &= \delta(q_1^1 = i) \\ \phi_{i,j,1}^{(Ver)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_t^1 = i) \cdot \delta(q_t^2 = j) \cdot \delta(f_{t-1} = 1) \\ \phi_{i,i',1}^{(Hor1)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_{t-1}^1 = i) \cdot \delta(q_t^1 = i') \cdot \delta(f_{t-1} = 1) \\ \phi_{j,j',i,0}^{(Hor2)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_{t-1}^2 = j) \cdot \delta(q_t^2 = j') \\ &\quad \cdot \delta(q_{t-1}^1 = i) \cdot \delta(f_{t-1} = 0) \\ \phi_{i,j,1}^{(End)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_t^1 = i) \cdot \delta(q_t^2 = j) \cdot \delta(f_t = 1) \\ \phi_{i,j}^{(Occ)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_t^1 = i) \cdot \delta(q_t^2 = j) \\ \Phi_{i,j}^{(M1)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_t^1 = i) \cdot \delta(q_t^2 = j) \cdot \mathbf{o}_t \\ \Phi_{i,j}^{(M2)}(Q^1, F, Q^2, O) &= \sum_{t=1}^T \delta(q_t^1 = i) \cdot \delta(q_t^2 = j) \cdot \mathbf{o}_t \mathbf{o}_t^T \end{aligned}$$

ここで、 $\phi_i^{(Prior)}$ は第 1 層の初期状態 q_1^1 が i のときに 1 をとる素性である. $\phi_{i,j,1}^{(Ver)}$ は第 1 層の状態 q_t^1 が i , かつ第 2 層の状態 q_t^2 が j , かつ指標変数 f_{t-1} が 1 となる回数を全時刻において数え上げる素性である. 他の素性の説明はスペースの都合上省略する.

† 広島市立大学大学院情報科学研究科 〒731-3194 広島市安佐南区大塚東 3-4-1
Email: takaaki@robotics.im.hiroshima-cu.ac.jp

2.2 パラメータ推定

学習データセット $\mathcal{D} = \{O^{(i)}, Q^{1(i)}\}_{i=1}^N$ が与えられたときのパラメータ Λ の推定を考える。ここでは HHMM と同様に最尤推定に基づくパラメータ学習を行う。HHMM との違いは、同時確率 $p(O, Q^1; \Lambda)$ を考えるのではなく、条件付き確率 $p(Q^1|O; \Lambda)$ を考える点にある。学習データにおける条件付き対数尤度は以下の式となる。

$$\mathcal{L}(\Lambda) = \sum_{i=1}^N \log p(Q^{1(i)}|O^{(i)}; \Lambda) \quad (2)$$

ここで、被推定パラメータ Λ の最適値 $\hat{\Lambda}$ は (2) 式の最大化問題で求まり、この際に必要な (2) 式の勾配は以下で表される。

$$\frac{\partial \mathcal{L}}{\partial \lambda_k} = \sum_{i=1}^N \left(\sum_{F, Q^2} \phi_k(Q^{1(i)}, F, Q^2, O^{(i)}) p(F, Q^2|Q^{1(i)}, O^{(i)}) - \sum_{Q^1} \sum_{F, Q^2} \phi_k(Q^1, F, Q^2, O^{(i)}) p(Q^1, F, Q^2|O^{(i)}) \right) \quad (3)$$

(3) 式の右辺は、確率分布 $p(F, Q^2|Q^{1(i)}, O^{(i)})$ による素性の期待値と、確率分布 $p(Q^1, F, Q^2|O^{(i)})$ による素性の期待値の差である。HHCRF は 2.1 節に定義した素性の形から、状態系列においてマルコフ過程を形成する。よって素性の期待値を計算する上では、前者の確率分布 $p(F, Q^2|Q^{1(i)}, O^{(i)})$ は $p(f_{t-1}, f_t, q_{t-1}^2, q_t^2|Q^{1(i)}, O^{(i)})$ と $p(f_t, q_t^2|Q^{1(i)}, O^{(i)})$ で十分であり、これは backward-forward-backward アルゴリズム [3] から求まる。後者の確率分布 $p(Q^1, F, Q^2|O^{(i)})$ も同様の理由により $p(q_{t-1}^1, q_t^1, f_{t-1}, f_t, q_{t-1}^2, q_t^2|O^{(i)})$ と $p(q_t^1, f_t, q_t^2|O^{(i)})$ で十分であり、これは forward-backward アルゴリズムから求まる。

2.3 ラベル推定アルゴリズム

周辺化ビタービアルゴリズムは、最適ラベル系列 \hat{Q}^1 を求めるアルゴリズムである。

$$\hat{Q}^1 = \operatorname{argmax}_{Q^1} \sum_{F, Q^2} p(Q^1, F, Q^2|O; \hat{\Lambda}) \quad (4)$$

これに対し、forward-backward アルゴリズムとビタービアルゴリズムは、最適ラベル系列 $[Q^1, F, Q^2]$ を求めるアルゴリズムである。

$$[\hat{Q}^1, \hat{F}, \hat{Q}^2] = \operatorname{argmax}_{Q^1, F, Q^2} p(Q^1, F, Q^2|O; \hat{\Lambda})$$

3 実験

脳波データ [4] のセグメンテーションに対して、HHMM と HHCRF の比較実験を行った。学習データには、主成分分析により 10 次元削減された長さ約 3500 の脳波データと、その脳波データに対応するラベル系列をそれぞれ 1 人当たり 3 本、合計 3 人分用意した。テストデータには同様の脳波データを 1 人当たり 1 本、合計 3 人分用意した。HHMM, HHCRF とともに 1 層目の状態数 3, 2 層目の状態数 5, 各時刻につき 1 つのガウス分布を持つ等価な 2 階層モデルとした。実験結果を表 1 に示す。表 1 には実験を 10 回繰り返した際の、3 つのラベル推定アルゴリズムごとのラベル正解率の平均値と、テストデータ O に対応

する被推定ラベル系列 \hat{Q} の対数尤度 $\log p(\hat{Q}|O; \hat{\Lambda})$ の平均値を示している。なお、HHCRF の初期パラメータには HHMM で学習したものを用いている。表 1 を見るように、どのラベル推定アルゴリズムにおいても HHCRF は HHMM よりもラベル平均正解率が高い。この結果から、時系列データのセグメンテーションにおいて、HHCRF の性能は HHMM よりも優れていることが確認できる。そして、本研究で提案した周辺化ビタービアルゴリズムは、当然ながら他のラベル推定アルゴリズムよりも高いラベル系列の尤度を持っている。

表 1 実験結果

アルゴリズム	モデル	平均正解率 (%)	平均対数尤度
forward-backward	HHCRF	76.69	3.4731e+04
	HHMM	64.98	8.4473e+04
ビタービ	HHCRF	65.46	1.4160e+05
	HHMM	59.76	1.3612e+05
周辺化ビタービ	HHCRF	65.27	1.4542e+05
	HHMM	59.76	1.4559e+05

4 まとめ

本研究では、階層 HMM に対応する識別モデル、階層隠れ CRF を提案した。階層隠れ CRF の推論アルゴリズムとして、上層のラベルのみからの学習アルゴリズム、下層の状態系列を周辺化し上層の最適系列を求める周辺化ビタービアルゴリズムの 2 つを考案した。そして、時系列データのセグメンテーションにおいて、識別モデルであるという違いから HHCRF の優位性を示した。

参考文献

- [1] David Pinto, Andrew MacCallum, Xing Wei, W. Bruce Croft, "Table Extraction Using Conditional Random Fields", In Proceedings of SIGIR '03 Conference, 235-242, 2003.
- [2] Kevin P. Murphy, Mark A. Paskin, "Linear Time Inference in Hierarchical HMMs", In Advances in Neural Information Processing Systems (NIPS), 2001.
- [3] Tobias Scheffer, Stefan Wrobel, "Active Learning of Partially Hidden Markov Models", Lecture Notes in Computer Science, 2001.
- [4] http://ida.first.fraunhofer.de/projects/bci/competition_iii/index.html
- [5] Asela Gunawardana, Milind Mahajan, Alex Acero, John C. Platt, "Hidden Conditional Random Fields for Phone Classification", In International Conference on Speech Communication and Technology, 2005.
- [6] Charles Sutton, Andrew McCallum, Khashayar Rohanimanesh, "Dynamic Conditional Random Fields: Factorized Probabilistic Models for Labeling and Segmenting Sequence Data", Journal of Machine Learning Research 8 (2007) 693-723.