

ニューラルセミマルコフCRFによる 論文アブストラクトの構造解析

山田 康輔¹ 平尾 努² 笹野 遼平¹ 武田 浩一¹ 永田 昌明²

¹ 名古屋大学 情報学研究科 ² NTT コミュニケーション科学基礎研究所

yamada.kosuke@c.mbox.nagoya-u.ac.jp

{tsutomu.hirao.kp,masaaki.nagata.et}@hco.ntt.co.jp

{sasano,takedasu}@i.nagoya-u.ac.jp

1 はじめに

論文アブストラクトの構造解析とは、アブストラクト中の各文に修辞役割ラベルを与えることでアブストラクトをいくつかの領域に分割するタスクである。図1に修辞役割ラベルが付与された生命医学文献データベース PubMed の論文アブストラクトの例を示す。この例における修辞役割ラベルは **BACKGROUND** (論文の背景), **OBJECTIVE** (論文の目的), **METHODS** (実験手法), **RESULTS** (実験結果), **CONCLUSIONS** (結論) の5つであり、文のかたまりに対しこれらのラベルを与えることでアブストラクトを5つの領域に分割している。こうした構造を解析することで、高度な検索、要約、流し読みなど論文からの知識獲得を支援することができる。

このタスクを従来の手法は文を単位とした系列ラベリング問題としてとらえている。つまり、個々の文に修辞役割ラベルと **B(egin)**, **I(nside)** タグを組み合わせて付与する問題として扱っており、現在の最高性能を達成したシステムもこの枠組みに基づいている [1, 2]。しかし、図1から明らかなように、基本的に同一の修辞役割ラベルは連続するため、文単位でラベルを推定するよりも領域そのものに対してラベルを与える問題として捉えた方がより効果的であると考えられる。

そこで、本研究ではアブストラクトをラベル付きスパン (連続する1文以上のまとまり) 系列の集合としてとらえ、ニューラルセミマルコフCRFを利用することで一番もっともらしい系列を選択する手法を提案する。従来の文単位で分類ラベルを推定する手法では局所的な情報しか利用できないのに対し、提案法を利用することでより広範囲な情報を利用できる。さらに、スパンに対して直接、分類ラベルを推定するため、領域分

Management of split-thickness skin graft donor sites: a randomized controlled trial of calcium alginate versus polyurethane film dressing.

Läuchli S¹, Hafner J, Ostheeren S, Mayer D, Barysch MJ, French LE.

Author information

¹ Dermatologische Klinik, Universitätsspital Zürich, Zürich, Switzerland.

Abstract

BACKGROUND: Split-thickness skin graft (STSG) donor sites sometimes cause more postoperative morbidity for patients than the wound covered with the graft. Yet, there is no consensus on which dressings are best suited to treat these donor sites.

OBJECTIVE: To evaluate two commonly used modern wound dressings in the postoperative healing of STSG donor sites in a prospective randomized controlled trial.

METHODS: 38 patients were randomly assigned to treatment of an STSG donor site with an alginate dressing or a polyurethane film dressing. The primary outcome measures were postoperative pain scores, secondary outcome variables were time to epithelialization, dressing changes and complications.

RESULTS: Postoperative pain on day 1 was significantly lower in the polyurethane film group (2.05 vs. 0.79, $p = 0.035$) as compared to the alginate group. This difference was not detected on day 5 (0.89 vs. 0.53, $p = 0.52$). Time to epithelialization did not differ significantly between the two dressing groups. There were more dressing changes in the polyurethane film group and problems with leakage.

CONCLUSION: Whereas film dressings resulted in initially lower pain scores, alginate dressings caused fewer additional dressing changes and less leakage.

図1: 修辞役割ラベルが付与された論文アブストラクトの例 (<https://www.ncbi.nlm.nih.gov/pubmed/24281776>)

割の性能向上が期待できる。文献 [3] のベンチマークデータを用いて提案法を評価したところ、個々の文に対するラベリングの F 値、スパンに対するラベリングの F 値とも世界最高性能を達成した。

2 関連研究

Dernoncourt ら [3] は、文字ベースの Bidirectional Long Short-Term Memory (Bi-LSTM) の隠れ状態ベクトルと事前学習済みモデルから獲得した単語ベクトルを利用して単語ベクトルを生成した後、それを入力とする Bi-LSTM を用いて文ベクトルを獲得した。そして、文ベクトルに対し、Conditional Random Fields (CRF) を用いてアブストラクトの構造を解析する手

法を提案した．文ベクトルを獲得するネットワーク層の上に最適なラベル系列を選択するための CRF を追加する枠組みはこのタスクの基本的な構成として広く用いられている．

Jin ら [1] の手法も Dernoncourt ら [3] の枠組みと基本的に同じであるが，文字ベースの Bi-LSTM は使わずに Attention を利用した Bidirectional Recurrent Neural Network (Bi-RNN) を用いることで，単語の重要度を考慮した文ベクトルを生成している．そして，文脈を考慮するために文ベクトル間にも Bi-LSTM を導入している．

Cohan ら [2] は汎用言語モデル Bidirectional Encoder Representations from Transformers (BERT) [4] を利用して文ベクトルを獲得している．Beltagy ら [5] による SciBERT の事前学習済みの重みを用いて，BERT の入力文の間に挿入する特殊トークン [SEP] の位置に対応した BERT の出力ベクトルを利用して分類ラベルを推定し，高い性能を達成している．

3 ニューラルセミマルコフ CRF

本研究では，文ではなく文のかたまりであるスパンに分類ラベルを付与することで論文アブストラクトの構造を解析する．この場合，入力となる文の系列 \mathbf{X} と出力であるラベル付きスパン系列 \mathbf{y} の長さは基本的には異なるため，セミマルコフ CRF を用いてモデル化し，スパンを特徴ベクトルと表現するため Bi-LSTM を利用する．

提案モデルの概要を図 2 に示す．このモデルは大きく 2 つのネットワーク層から構築される．1 つは文ベクトルからスパン特徴量ベクトルを生成する層と，もう 1 つはスパン特徴量ベクトルを利用したセミマルコフ CRF によってスパン系列のラベリングを学習する層である．

本研究におけるニューラルセミマルコフ CRF モデルは文の系列 \mathbf{X} に対するラベル付きスパン系列 \mathbf{y} の条件付き確率を $P(\mathbf{y}|\mathbf{X})$ として正解のラベル付きスパン系列 \mathbf{y}^* に対する対数尤度関数を最大化するようパラメタを学習する．

$P(\mathbf{y}|\mathbf{X})$ は式 1 で表されるようにスパン系列のスコアにソフトマックス関数を適用することで得られる．

$$P(\mathbf{y}|\mathbf{X}) = \frac{\exp(\text{score}(\mathbf{X}, \mathbf{y}))}{\sum_{\mathbf{y}' \in \mathbf{Y}} \exp(\text{score}(\mathbf{X}, \mathbf{y}'))} \quad (1)$$

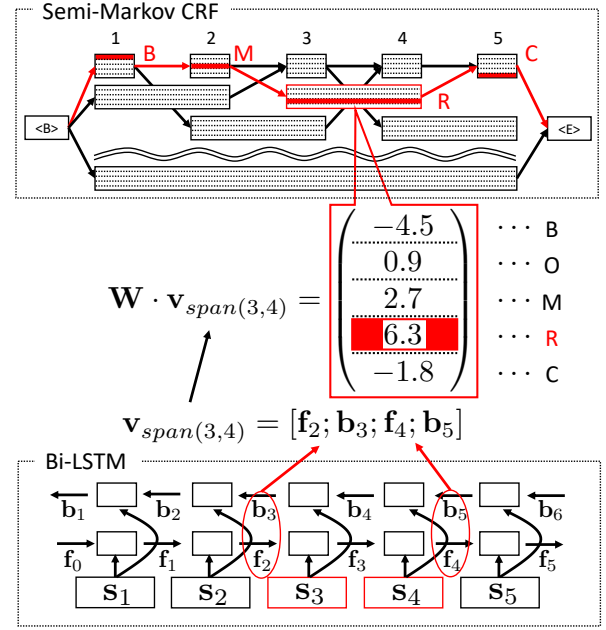


図 2: ニューラルセミマルコフ CRF モデルの概要．この例では，修辞役割ラベルが 5 タイプ {B, O, M, R, C}，論文アブストラクトは 5 文から成り，各文のラベルが [B, M, R, R, C] となる．

\mathbf{Y} は \mathbf{X} から得られるすべてのラベル付きスパン系列の集合であり，ラベル付きスパン系列 \mathbf{y} をその長さを m として $[y_1, y_2, \dots, y_m]$ とする．そして， k 番目のスパン y_k をその開始位置 i_k ，終了位置 j_k ，ラベル ℓ_k を用いて (i_k, j_k, ℓ_k) とする．

ラベル付きスパン系列のスコア $\text{score}(\mathbf{X}, \mathbf{y})$ は，以下の式で定義する．

$$\text{score}(\mathbf{X}, \mathbf{y}) = \sum_{k=1}^m e_{(i_k, j_k, \ell_k)} + \sum_{k=1}^m T_{\ell_k, \ell_{k+1}} \quad (2)$$

式 (2) の右辺第 1 項はスパンに対して分類ラベルを与えるスコア，第 2 項は分類ラベル間の遷移スコアである．

3.1 スパン特徴量ベクトル

式 (2) を計算するため，任意のスパンをベクトルとして表現する．まず，後に述べる方法で任意の文をベクトル \mathbf{s}_i として表現し，前向き LSTM，後ろ向き LSTM を用いてそれぞれの隠れ状態ベクトル \mathbf{f}_i ， \mathbf{b}_i を以下の式で得る．それぞれのベクトルの大きさは d_h である．

$$\begin{aligned} \mathbf{f}_i &= \overrightarrow{\text{LSTM}}(\mathbf{f}_{i-1}, \mathbf{s}_i), \\ \mathbf{b}_i &= \overleftarrow{\text{LSTM}}(\mathbf{b}_{i+1}, \mathbf{s}_i), \end{aligned} \quad (3)$$

そして、位置 i から位置 j までのスパンを $span(i, j)$ と表し、スパンの特徴量ベクトル $\mathbf{v}_{span(i, j)}$ を式 4 により獲得する。

$$\mathbf{v}_{span(i, j)} = [\mathbf{f}_{i-1}; \mathbf{b}_i; \mathbf{f}_j; \mathbf{b}_{j+1}] \quad (4)$$

次に、 $\mathbf{v}_{span(i, j)}$ を用いてスパンに修辭役割ラベルを与えるスコア $e_{(i, j, \ell)}$ を重み行列 $\mathbf{W} \in \mathbb{R}^{label_size \times 4d_h}$ を用いて以下の式で獲得する。

$$\mathbf{e}_{(i, j, \ell)} = \mathbf{W} \cdot \mathbf{v}_{span(i, j)} \quad (5)$$

また、遷移スコア \mathbf{T} は同じ修辭役割ラベル間の遷移は起こらないものとし、 ℓ_0 と ℓ_{m+1} はラベル系列の先頭と末尾を示すラベル $\langle B \rangle$ と $\langle E \rangle$ とする。

4 実験

4.1 データセット

実験には、Dernoncourt ら [3] の PubMed-RCT データセット¹を利用した²。生命医学文献データベース PubMed 内のランダム化比較試験 (Randomized Controlled Trials) に関する論文のアブストラクトの各文に修辭役割ラベルが付与されたデータセットであり、そのラベルは先述の例と同じく **BACKGROUND**, **OBJECTIVE**, **METHODS**, **RESULTS**, **CONCLUSIONS** の 5 つである。アブストラクトは全体で 20000 件あり、うち学習データとして 15000 件、開発データ、テストデータがそれぞれ 2500 件である。なお、実験に用いたすべての手法のハイパーパラメタは開発データを用いて決定した。

表 1 にデータセット内の文単位のスパン単位とラベル列の出現頻度上位 5 つを示す。それぞれ合計 5203, 88 種ある。文単位の集計結果では O が先頭に 1 つ出現し、M と R が複数続き、C が 1 つで続くデータが多い。スパン単位の集計結果では出現頻度上位 5 つで 85% 以上を占めており、B または O で始まり M→R→C と続く、あるいは B→O のあとに M→R→C と続くものが多い。さらに、C のあとに B が出現する場合も少なからず存在する。

4.2 実験設定

ニューラルセミマルコフ CRF モデルの入力となる文のベクトル表現として、本研究では 2 つの事前学習

表 1: データセット内の文単位のスパン単位とラベル列の出現頻度上位 5 つ。ラベルの後ろの数字はラベルが続いた個数を表す。

文単位	データ数	スパン単位	データ数
O1M3R3C1	169	OMRC	8395
O1M4R3C1	155	BMRC	4376
O1M3R4C1	131	BOMRC	2165
O1M5R3C1	122	BMRCB	1324
O1M4R4C1	122	OMRCB	850
5203 種	20000	88 種	20000

済みモデルを利用して、4 種類のバリエーションを試した。1 つ目は Peters ら [6] による PubMed 内の論文アブストラクトを利用して事前学習された Embeddings from Language Models (ELMo) から獲得する方法である³。ELMo に対して、文単位で入力し、出力として各単語ベクトルを獲得できる。その単語ベクトルを平均プーリングすることで、文ベクトル (ELMo-AVE) とする。文ベクトルの次元数は 1024 である。

2 つ目は Peng ら [7] による PubMed 内の論文アブストラクトを利用して事前学習された BERT から獲得する方法である⁴。BERT に対して、ラベルが付与される文単位で入力する。このとき、文の先頭にタグ [CLS]、文の末尾にタグ [SEP] を挿入し、入力している。そして、最後から 2 番目の層のタグ [CLS] のベクトルと各単語のベクトルを獲得した。BERT から 3 種類の文ベクトルを獲得している。1 つ目はタグ [CLS] のベクトルをそのまま文ベクトルとして扱ったもの (BERT-CLS)、2 つ目は単語ベクトルを平均プーリングした文ベクトル (BERT-AVE)、3 つ目は単語ベクトルを最大プーリングした文ベクトル (BERT-MAX) である。BERT による文ベクトルの次元数は 728 である。

また、比較手法として入力の文ベクトルとして BERT-CLS を用いた Bi-LSTM+CRF [8]、Jin らの手法 [1]、Cohan らの手法 [2] を用い、評価指標には sentence-F1 と span-F1 の 2 つを利用した。sentence-F1 は文単位でラベルの一致を micro-F1 で評価し、span-F1 はスパン単位でラベルの一致を micro-F1 で評価する。

¹<https://github.com/Franck-Dernoncourt/pubmed-rct>

²数字を@に置き換えていない生データを利用している。

³<https://allennlp.org/elmo>

⁴https://github.com/ncbi-nlp/ncbi_bluebert

表 2: Bi-LSTM+CRF, 先行研究のモデルと比較した実験結果. カッコ内は利用した文ベクトルを示す.

	sentence-F1	span-F1
Bi-LSTM+CRF	91.8	81.2
Jin ら	92.8	82.9
Cohan ら	92.9	82.2
(ELMo-AVE)	88.4	70.5
提案モデル (BERT-CLS)	93.1	84.3
(BERT-AVE)	92.7	83.8
(BERT-MAX)	90.4	76.7

表 3: ラベルごとの span-F1

	Jin ら	Cohan ら	提案モデル
BACKGROUND	73.8	70.6	74.7
OBJECTIVE	73.8	70.8	73.8
METHODS	86.7	86.3	88.5
RESULTS	83.1	83.9	85.8
CONCLUSIONS	90.8	92.0	91.9

4.3 実験結果

表 2 に実験結果を示す. 文ベクトル (BERT-CLS) を利用した本研究の提案モデルが sentence-F1, span-F1 の共に世界最高性能を達成した. 特に span-F1 による評価において, F 値の向上が顕著であり提案モデルが他のモデルよりラベル付きスパンを正しく推定できていることがわかる. Bi-LSTM+CRF と提案モデルは入力として同じ文ベクトル BERT-CLS を利用しており, これらのモデルの違いは可変長のスパンを扱えるかどうかである. 実験結果からは, スパンを考慮したモデルにすることで, F 値の向上していることが確認できるので論文アブストラクトの構造解析にニューラルセミマルコフ CRF を導入する利点がよくわかる. また, 提案モデルにおける文ベクトルの比較では, ELMo を利用して獲得したベクトルより BERT を利用して獲得したベクトルの方が全体的により高い F 値を達成することが確認できた.

表 3 にラベルごとの span-F1 を示す. すべてのラベルにおいて従来法と同等以上の span-F1 の F 値を達成した. 特に, METHODS と RESULTS では, 提案モデルが他モデルより良い. METHODS, RESULTS スパンは他のラベルのスパンより長いという特徴があり, スパンにラベル付けを行う提案モデルが有効であったと考える.

5 おわりに

本研究では, 論文アブストラクトの構造解析において, アブストラクト中の個々の文に修辭役割ラベルを与え, 全体の構造を推定するのではなくアブストラクト中の文のかたまりであるスパンに直接修辭役割ラベルを与えるニューラルセミマルコフ CRF モデルを提案した. ベンチマークデータを用いて提案法を評価したところ, sentence-F1, span-F1 とともに世界最高性能を達成した. 特に, span-F1 の向上が顕著であり, スパンに修辭役割ラベルを直接与えることの有効性がわかった. 各ラベルの span-F1 においても, 提案モデルの有用性を確認することができ, METHODS, RESULTS といった比較的長いスパンではスコア向上が顕著であった. 提案法は文のベクトル表現としては BERT を利用しただけであるが, これを精緻化することでさらに性能を改善できるのではないかと考えている.

参考文献

- [1] Di Jin and Peter Szolovits. Hierarchical neural networks for sequential sentence classification in medical scientific abstracts. In *Proc. of EMNLP'18*, pp. 3100–3109, 2018.
- [2] Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Dan Weld. Pretrained language models for sequential sentence classification. In *Proc. of EMNLP-IJCNLP'19*, pp. 3691–3697, 2019.
- [3] Franck Dernoncourt and Ji Young Lee. Pubmed 200k rct: a dataset for sequential sentence classification in medical abstracts. In *Proc. of IJCNLP'17*, pp. 308–313, 2017.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of NAACL-HLT'19*, pp. 4171–4186, 2019.
- [5] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pre-trained language model for scientific text. In *Proc. of EMNLP-IJCNLP'19*, pp. 3613–3618, 2019.
- [6] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proc. of NAACL-HLT'18*, pp. 2227–2237, 2018.
- [7] Yifan Peng, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: An evaluation of bert and elmo on ten benchmarking datasets. In *Proc. of BioNLP'19*, pp. 58–65, 2019.
- [8] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging, 2015. In Arxiv preprint <https://arxiv.org/abs/1508.01991>.