

アテンションCNNによる薬物間相互作用抽出

Extraction of drug-drug interactions with attention CNNs

浅田真生 三輪誠 佐々木裕
Masaki Asada Makoto Miwa Yutaka Sasaki

豊田工業大学
Toyota Technological Institute

Convolutional neural networks (CNNs), automated feature learning methods, are shown to have a great potential in the drug-drug interactions (DDI) extraction task. The CNN with attention mechanism, which emphasizes important words in the sentence of a entity pair, is also reported to be effective for the task of relation classification. In this research, we apply the CNN with attention mechanism to the task of DDI extraction. We evaluated the model on the Semeval-2013 DDI extraction challenge dataset. As a result, the model archived an F-score of 68.08%, which outperformed the CNN-based method without attention mechanism by 1.05 percent points.

1. はじめに

患者に薬物を併用投与する場合、薬品の本来の作用が增強・減弱することや、副作用が起きることがあり、このことを薬物間相互作用とよぶ。薬物間相互作用をまとめたデータベースが医療専門家のために多く提供されている [Law 14] が、薬学文書は急速に増大しているため、新たに発見された相互作用への対応が遅い。更新の早いデータベースの構築を可能とするため、生体医学に関する文章から薬物間相互作用を抽出する薬物間相互作用抽出の研究は広く行われている [Kim 15, Chowdhury 15].

文書からの薬物間相互作用抽出に対しては、近年、ニューラルネットワークを用いた手法が注目されている。ニューラルネットワークの中でも畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) は、計算が速いという利点を持ち、CNN を用いた薬物間相互作用抽出 [Liu 16] は高い性能を示している。

また、Wang らは薬学系文書ではない一般的な文書の関係分類のタスク [Hendrickx 10] において、入力文の中で関係分類に重要である部分、特に関係の種類を求めるターゲットの2単語との関係が大きい部分に注目するアテンションCNN [Wang 16] を提案し、高い性能を示している。

本研究では、薬物間相互作用抽出のタスクで、アテンションを用いたCNNによる手法を学習・評価し、アテンションが一般的な文書を対象としたタスクだけでなく、薬学文書を対象とした関係抽出である薬物間相互作用抽出においても有用であるかを検証することを目的とする。

2. 関連研究

2.1 CNNによる薬物間相互作用抽出

Liu らは薬物間相互作用抽出の精度向上を目的とし、CNNを用いた手法 [Liu 16] を提案し、人手による特徴付与を必要とする手法と比較して高い性能を示している。入力文 $S = w_1, w_2, \dots, w_n$ と薬物の単語 e_1, e_2 が与えられたときに薬物のペアが取る関係の確率 P を求める。

2.1.1 単語ベクトル

単語 w_i ($i = 1, 2, \dots, n$) の単語ベクトル \mathbf{w}_i^w を $d^w \times |V|$ 次元の単語埋め込み行列 \mathbf{W}^{wrd} を用いて式 (1) のように求める。

$$\mathbf{w}_i^w = \mathbf{W}^{wrd} \mathbf{v}_i^w \quad (1)$$

ここで d^w は単語ベクトルの次元数、 V は入力の語彙である。 \mathbf{W}^{wrd} のそれぞれの列 \mathbf{W}_j^{wrd} は、語彙の j 番目の単語のベクトル表現で表される。 \mathbf{v}_i^w は長さ $|V|$ のベクトルで、単語 w_i が対応するインデックスの要素が1、それ以外の要素が0である。

2.1.2 単語位置ベクトル

単語位置ベクトルは、ターゲットの単語との距離についての情報をベクトルで表現したものである。単語位置ベクトルは、現在の単語とターゲットの単語との相対的な距離から求める。例文 “The [car] left the [plant]” が与えられたときに、単語 “left” の “car”, “plant” に対する相対的距離は $(-1, 2)$ である。それぞれの相対的距離に対して次元数 d^p のベクトルが割り当てられる。ただしベクトルの初期値はランダムに決定する。単語 w_i ($i = 1, 2, \dots, n$) の、ターゲットの単語 e_1, e_2 に対する単語位置ベクトルを $\mathbf{w}_{i,1}^p, \mathbf{w}_{i,2}^p$ とする。

2.1.3 畳み込み

単語ベクトルと単語位置ベクトルを連結し、 w_i を式 (2) のように表現する。

$$\mathbf{w}_i^W = [(\mathbf{w}_i^w)^T, (\mathbf{w}_{i,1}^p)^T, (\mathbf{w}_{i,2}^p)^T]^T \quad (2)$$

また、フィルターサイズの異なる L 個の畳み込みフィルターを用意し、フィルターサイズの集合を式 (3) のように表す。

$$\mathbf{k} = \{k_1, k_2, \dots, k_l, \dots, k_L\} \quad (3)$$

ここで l 番目のフィルターのサイズは k_l である。Liu らはフィルターサイズとして $\{3, 4, 5\}$ を用いている。単語 w_i とその後 $\frac{k_l-1}{2}$ 単語のベクトルを連結し、式 (4) のように表現する。

$$\mathbf{z}_{i,l} = [(\mathbf{w}_{i-(k_l-1)/2}^W)^T, \dots, (\mathbf{w}_{i-(k_l+1)/2}^W)^T]^T \quad (4)$$

畳み込みの重み行列を $\mathbf{W}^{conv} \in \mathbb{R}^{d^c \times (d^w + 2d^p) \times k}$ とし、 \mathbf{W}^{conv} の j 番目の列を $\mathbf{W}_j^{conv} \in \mathbb{R}^{(d^w + 2d^p) \times k}$ とする。 d^c は畳み込みのフィルタ数である。 $\mathbf{z}_{i,l}$ に畳み込み処理を行い、以下の結果を得る。

$$m_{i,j,l} = \tanh(\mathbf{W}_j^{conv} \odot \mathbf{z}_{i,l} + b) \quad (5)$$

ここで \odot は要素積、 b はバイアス項である。

2.1.4 max プーリングおよびソフトマックス

全単語の畳み込みの出力のうち、値が最大であるものを選択する max プーリングにより以下の結果を得る。

$$c_{j,l} = \max_i m_{i,j,l} \quad (6)$$

$$\mathbf{C}_l = [c_{1,l}, c_{2,l}, \dots, c_{d_c,l}] \quad (7)$$

$$\mathbf{C} = [(\mathbf{C}_1)^T, (\mathbf{C}_2)^T, \dots, (\mathbf{C}_l)^T, \dots, (\mathbf{C}_L)^T]^T \quad (8)$$

重み行列 $\mathbf{W} \in \mathbb{R}^{o \times d^c}$ を用いて \mathbf{C} を変換し、

$$\mathbf{I} = \mathbf{W}\mathbf{C} \quad (9)$$

とする。ここで o は分類する関係の総数、 \mathbf{I} は $\mathbf{I} = [i_1, i_2, \dots, i_o]$ と表される。得られた \mathbf{I} にソフトマックス関数を施し、取りうる関係の確率 \mathbf{P} を以下のように求める。

$$\mathbf{P} = [p_1, p_2, \dots, p_o] \quad (10)$$

$$p_j = \frac{\exp i_j}{\sum_{k=1}^o \exp i_k} \quad (11)$$

正解ラベルを \mathbf{y} とし、式 (12) に示す目的関数を最小にするように各種パラメータを更新する。

$$L = - \sum \mathbf{y} \log \mathbf{P} \quad (12)$$

2.2 アテンション CNN による一般名詞関係分類

Wang らは、一般文書関係分類 [Hendrickx 10] のさらなる精度向上を目的とし、ターゲットの単語との関係が大きい単語を強調するアテンションを用いた CNN [Wang 16] による手法を提案し、高い性能を示している。単語 w_i に対するアテンション α_i を以下のように求める。はじめに、ターゲットの単語 e_j に対して、ベクトル \mathbf{a}^j を式 (13) のように定義する。

$$\mathbf{a}_i^j = \mathbf{w}_{e_j}^M \cdot \mathbf{w}_i^M \quad (13)$$

ここで \mathbf{a}_i^j は \mathbf{a}^j の i 番目の要素、 $\mathbf{w}_{e_j}^M$ はターゲットの単語 e_j のベクトル表現である。対角行列 \mathbf{A}^j にソフトマックス関数を施して、アテンション α_i を式 (15) のように表す。

$$\alpha_i^j = \frac{\exp \mathbf{a}_i^j}{\sum_{k=1}^n \exp \mathbf{a}_k^j} \quad (14)$$

$$\alpha_i = \frac{\alpha_i^1 + \alpha_i^2}{2} \quad (15)$$

得られたアテンション α を畳み込み処理の直前に施す。 \mathbf{z}_i にアテンションを施したものを \mathbf{z}_i^α とし式 (16) に示す。

$$\mathbf{z}_i^\alpha = \mathbf{z}_i \alpha_i \quad (16)$$

3. 提案手法

アテンションを用いた CNN が薬物間相互作用抽出において有用であるかを検証するために、薬物間相互作用抽出のタスクを、アテンション CNN により学習・評価する。提案手法では、薬物を含んだ文章が与えられたときの相互作用抽出を以下のようにして行う。ただし、薬物を指示する用語はすでに特定されているとする。

薬物を n 個含む文が与えられとき、相互作用を持つ可能性のあるペアが nC_2 個作成される。先行研究 [Liu 16] と同じ条件で評価を行うために、以下に示す前処理を行う。

- 文中の 2 つのターゲットの薬物の単語をそれぞれ "DRUG1", "DRUG2" で置き換える。
- 文中のターゲット以外の薬物の単語を "DRUGOTHER" で置き換える。

この前処理を行うことにより、作成する相互作用抽出器が薬物の表層のみに依存することを防ぎ、文脈全体の情報を用いた相互作用抽出を行うことができるようになる。

前処理によって得られた入力文に対して、単語分割を行い、全ての大文字を小文字に変換する。分割されたそれぞれの単語を、単語ベクトルおよび単語位置ベクトルを用いて数値ベクトル化し、入力文を行列で表現する。単語 w_i ($i = 1, 2, \dots, N$) の単語ベクトル \mathbf{w}_i^d を $d^w \times |V|$ 次元の単語埋め込み行列 \mathbf{W}^{wrd} を用いて式 (17) のように求める。

$$\mathbf{w}_i^d = \mathbf{W}^{wrd} \mathbf{v}_i^w \quad (17)$$

ここで d^w は単語ベクトルの次元数、 V は単語ベクトルの事前学習によって得られた語彙で、 \mathbf{v}_i^w により \mathbf{W}^{wrd} から単語 w_i が対応する単語ベクトルを抜き出す。単語埋め込み行列 \mathbf{W}^{wrd} は学習により更新される。また、単語 w_i ($i = 1, 2, \dots, n$) の、ターゲットの単語 e_1, e_2 に対する単語位置ベクトルを $\mathbf{w}_{i,1}^p, \mathbf{w}_{i,2}^p$ とする。単語ベクトルおよび単語位置ベクトルは正規化を行う。

単語ベクトルおよび単語位置ベクトルを用いてアテンションを算出し、CNN による学習を行う。アテンションを用いた CNN モデルの概要を図 1 に示す。単語ベクトル \mathbf{w}_i^d と単語位置ベクトル $\mathbf{w}_{i,1}^p, \mathbf{w}_{i,2}^p$ を連結し、単語 w_i を式 (18) のようにベクトルで表現する。

$$\mathbf{w}_i^M = [(\mathbf{w}_i^d)^T, (\mathbf{w}_{i,1}^p)^T, (\mathbf{w}_{i,2}^p)^T]^T \quad (18)$$

また、フィルターサイズが k_l ($l = 1, \dots, L$) であるとき、単語 w_i とその前後 $\frac{k_l-1}{2}$ 単語のベクトルを連結し、式 (19) のように表現する。

$$\mathbf{z}_{i,l} = [(\mathbf{w}_{i-(k_l-1)/2}^M)^T, \dots, (\mathbf{w}_{i+(k_l+1)/2}^M)^T]^T \quad (19)$$

提案手法では、アテンションを以下のようにターゲットの薬物の単語を無視して計算する。ただし入力文の e_1, e_2 番目の単語が薬物の単語 1、薬物の単語 2 を示すとする。

$$\mathbf{a}_i^j = \mathbf{w}_{e_j}^M \mathbf{w}_i^M \quad (20)$$

$$\alpha_i^j = \begin{cases} \frac{\exp \mathbf{a}_i^j}{\sum_{1 \leq k \leq n, k \neq e_1, e_2} \exp \mathbf{a}_k^j} & (i \neq e_1, e_2) \\ \alpha_{drug} & (otherwise) \end{cases} \quad (21)$$

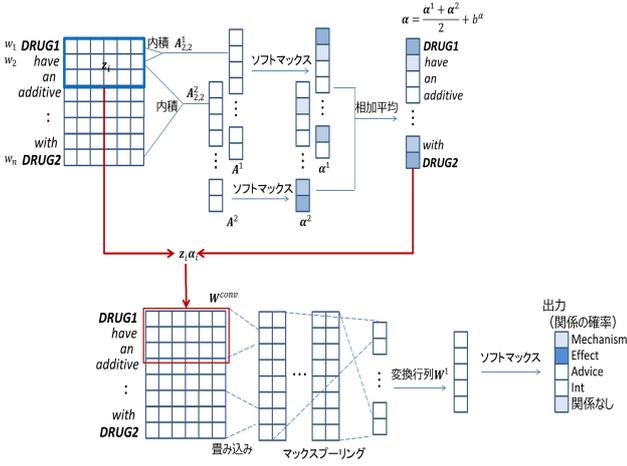


図 1: アテンションを用いた CNN

$$\alpha_i = \frac{\alpha_i^1 + \alpha_i^2}{2} + b^\alpha \quad (22)$$

ここで、 b^α はバイアス項で、 α^{drug} は学習により決定する。畳み込みの重み行列を $\mathbf{W}^{conv} \in \mathbb{R}^{d^c \times d^w \times k}$ とし、 \mathbf{W}^{conv} の j 番目の列を $\mathbf{W}_j^{conv} \in \mathbb{R}^{d^w \times k}$ とする。 d^c は畳み込みのフィルタ数である。 $\mathbf{z}_{i,l}$ にアテンション α_i を掛け、畳み込み処理を行い、以下の結果を得る。

$$m_{i,j,l} = f(\mathbf{W}_j^{conv} \odot \mathbf{z}_{i,l} \alpha_i + b) \quad (23)$$

ここで \odot は要素積、 b はバイアス項、 f は ReLU 関数で。

$$f(x) = \begin{cases} x & (x \geq 0) \\ 0 & (otherwise) \end{cases} \quad (24)$$

を満たす。

畳み込みにより得られた結果に対して max プーリングを行い以下の結果を得る。

$$c_{j,l} = \max_i m_{i,j,l} \quad (25)$$

$$\mathbf{C}_l = [c_{1,l}, c_{2,l}, \dots, c_{d_c,l}] \quad (26)$$

$$\mathbf{C} = [(\mathbf{C}_1)^T, (\mathbf{C}_2)^T, \dots, (\mathbf{C}_l)^T, \dots, (\mathbf{C}_L)^T]^T \quad (27)$$

重み行列 $\mathbf{W} \in \mathbb{R}^{o \times d^c}$ を用いて \mathbf{C} を変換し、

$$\mathbf{I} = \mathbf{W}\mathbf{C} \quad (28)$$

とする。ここで o は分類する関係の総数である。 $\mathbf{I} = [i_1, i_2, \dots, i_o]$ と表されるとき、 \mathbf{I} にソフトマックス関数を施し、取りうる関係の確率 \mathbf{P} を以下のように求める。

$$\mathbf{P} = [p_1, p_2, \dots, p_o] \quad (29)$$

$$p_j = \frac{\exp i_j}{\sum_{k=1}^o \exp i_k} \quad (30)$$

正解ラベルを \mathbf{y} とし、式 (31) に目的関数 L を示す。

$$L = - \sum \mathbf{y} \log \mathbf{P} \quad (31)$$

また、過学習を防ぐために L2 正則化を行う。式 (32) に L2 正則化を適用した目的関数 L' を示す。

$$L' = L + \lambda (\|\mathbf{W}^{word}\|_F^2 + \|\mathbf{W}^{conv}\|_F^2 + \|\mathbf{W}^1\|_F^2) \quad (32)$$

L' を最小にするように各パラメータを更新する。最適化はオンライン学習により行う。学習データをランダムに並び替えて分割し、ミニバッチファイルを作成する、その際に、アンダーサンプリングを行う。アンダーサンプリングとは、学習データが正例と比較して負例を多く含む場合に、データの偏りを軽減するために負例を減らすことである。最適化アルゴリズムとして Adam [Kingma 15] を使用する。また、バッチ正規化 [Ioffe 15] を行う。

4. 結果と考察

4.1 実験設定

4.1.1 データセット

使用するデータセットは、Semeval-2013 Task 9 データセット [Segura-Bedmar 13] である。このデータセットは、薬物の単語を含んだ文章から構成され、どの単語が薬物であるかはあらかじめ特定されている。データセットは MEDLINE のアブストラクトと DrugBank データベースの 2 つからなる。このデータセットは以下に示す 4 種類の相互作用が注釈付けされており、本実験では薬物のペアの相互作用の有無および、相互作用を持つ場合は 4 種類のうちのどの相互作用を持つかを求める。

Mechanism: 文章が薬力学的機構を表す。

Effect: 文章が薬物の効果を表す。

Advice: 文章が 2 つの薬物を併用する際の推奨を表す。

Int: 文章が 2 つの薬物が相互作用を持つことのみを表す。

データセットの内訳を表 1 に示す。表 1 に示すように相互作用を持つペア (正例) 数に比べて相互作用を持たないペア (負例) 数が大きい。

4.1.2 各種パラメータ

本実験で使用したハイパーパラメータを表 2 に示す。本実験は、学習データを分割し開発データを作成し、開発データを用いてパラメータチューニングを行った。単語分割器として GENIA tagger を使用し、単語ベクトルの事前学習は Skip-gram [Mikolov 13] によって行った。事前学習に用いたコーパスは PubMed2014 のデータで、語彙数は 1,630,978 である。"DRUG1", "DRUG2" として置き換えられたターゲットの薬物の単語ベクトルの初期値は、ともに "drug" という単語と同じ値を用いる。ミニバッチファイルを作成するときに、負例が選ばれる確率を 1/4 まで小さくし、データの偏りを防ぐ。

4.2 結果

アテンションを用いない CNN による手法とアテンションを用いた CNN による手法を F 値で評価したものを表 3 に示

表 1: 実験データセットの内訳

	学習データ	開発データ	評価データ
文数	6,976	1,404	1,299
ペア数	27,792	4,998	5,716
相互作用を持つペア数	4,021	751	979
相互作用を持たないペア数	23,771	4,247	4,737
Mechanism	1,319	232	302
Effect	1,687	339	360
Advice	826	132	221
Int	189	48	96

す。アテンションを用いない場合と比較して、全体の F 値が 1.05pp 上昇した。表 4 に CNN モデル [Liu 16] と提案手法のモデルを、負例のフィルタリングの有無、F 値で比較したものを示す。負例のフィルタリングとは、ルールベースで作成した分類器により、関係を持たないペア（負例）の一部を省いて学習を行うことである。負例のフィルタリングを行わない CNN モデルと比較して、提案手法が高い F 値を示した。一方で、負例のフィルタリングを行った CNN モデルと提案手法のモデルを比較すると、提案手法の F 値が下回ることがわかった。

また、アテンションの追加により正解できるようになった問題文のアテンションを可視化したものを図 2 に示す。例文は、DRUG1 と DRUG2 が “Effect” という関係を持つことを示している。アテンションにより、“increase”, “effect”, “of” といった単語が強調されていることが分かった。

4.3 考察

注意機構を用いた CNN による手法が、注意機構を用いない CNN による手法と比較して、全体の F 値が 1.05pp したことから、注意機構が薬物相互作用抽出において有効であると考えられる。また、先行研究 [Liu 16] の報告の通り、負例のフィルタリングが相互作用抽出において有効である。しかし、負例のフィルタリングは人手による特徴付与を必要としコストが大きく、分類器の汎用性が失われる可能性があるため、負例のフィルタリングを行わずにより高い性能を示すモデルの提案が必要であると考えられる。

表 2: 各種パラメータ

パラメータ	値
単語ベクトルの次元数	200
単語位置ベクトルの次元数	20
畳み込みのフィルターサイズ	{3,4,5}
畳み込みのフィルター数	100
学習率	0.001
ミニバッチサイズ	100
L2 正則化の係数	0.0001

表 3: アテンションの有無における F 値の比較

アテンション	無			有		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
Mechanism	68.90	64.57	66.67	71.96	64.57	68.06
Effect	68.73	61.67	65.01	71.15	60.28	65.26
Advice	80.93	71.04	75.66	79.23	74.21	76.64
Int	77.78	36.46	49.65	83.72	37.50	51.80
All	72.07	62.65	67.03	74.09	62.96	68.08

表 4: 各手法の F 値の比較

手法	負例のフィルタリング	F 値 (%)
CNN[Liu 16]	有	69.75
CNN[Liu 16]	無	67.01
提案手法	無	68.08

DRUGOTHER	and	DRUG1	increase	the
effects	of	DRUG2	.	

図 2: アテンションの可視化

5. まとめ

本研究では、一般文書の関係分類のタスクで高い精度を示しているアテンション CNN が、薬物間相互作用抽出のタスクに対して有用であるかを検証することを目的に、アテンション CNN を薬物間相互作用抽出のタスクで学習・評価し、アテンションを用いない場合と比較して高い F 値を得た。この結果より、アテンションが薬物間相互作用抽出のタスクにおいても有効であることを示した。今後の課題は、負例のフィルタリングを行わずにより高い性能を示すモデルの提案である。

参考文献

- [Law 14] Vivian Law et al.: DrugBank 4.0: shedding new light on drug metabolism (2014), *Nucleic Acids Research*.
- [Kim 15] Sun Kim et al.: Extracting drug-drug interactions from literature using a rich feature-based linear kernel approach (2015), *Journal of biomedical informatics*.
- [Chowdhury 15] Md Faisal Mahbub Chowdhury et al.: FBK-irst : A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information (2013), In *Proceedings of the 7th International Workshop on Semantic Evaluation*.
- [Liu 16] Shengyu Liu et al.: Drug-drug interaction extraction via convolutional neural networks (2016), *Computational and mathematical methods in medicine*.
- [Wang 16] Linlin Wang et al.: Relation classification via multi-level attention cnns (2016), In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.
- [Hendrickx 10] Iris Hendrickx et al.: Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals (2010), In *Proceedings of the 5th International Workshop on Semantic Evaluation*.
- [Kingma 15] Diederik P. Kingma et al.: Adam: A Method for Stochastic Optimization (2015), In *Proceedings of the 3rd International Conference for Learning Representations*.
- [Ioffe 15] Sergey Ioffe et al.: Batch normalization: Accelerating deep network training by reducing internal covariate shift (2015), In *COLING*.
- [Segura-Bedmar 13] Isabel Segura-Bedmar et al.: SemEval-2013 Task 9: Extraction of DrugDrug Interactions from Biomedical Texts (2013), In *Proceedings of the 7th International Workshop on Semantic Evaluation*.
- [Mikolov 13] Tomas Mikolov et al.: Distributed Representations of Words and Phrases and their Compositionality (2013), In *Proceedings of the 26th Annual Conference on Advances in Neural Information Processing Systems*.