BERT と GAT を用いた金融テキストにおける因果関係を含む文の判定

小林 涼太郎 [1]　坂地 泰紀 [2]　和泉 潔 [2]
[1] 東京大学工学部 [2] 東京大学大学院工学系研究科
b2022rkobayashi@socsim.org
{sakaji,izumi}@sys.t.u-tokyo.ac.jp

概要

金融分野に関わる大量のテキストデータを解析し，人が認知する原因-結果関係についての記述を自動的に抽出することで，経済事象の要因列挙による投資判断の支援や，イベントの波及効果分析が可能となる．テキストから因果関係を抽出するタスクにおいては，因果関係の存在を示す手がかりとなる表現への注目が有効であることが知られている．しかしながら，手がかり表現が因果関係以外の意味を持つ場合もあり，それを取り除くために，文に因果関係が含まれているか否かを判定する手法が必要である．本研究では，金融 BERT モデルと，入力文の依存構造に対して適用される GAT の組み合わせによる，新しい因果関係判定手法を提案する．

1　はじめに

近年，Web 上で入手可能なテキストデータは急速に増大しており，自然言語処理技術を用いた，膨大なテキストからの情報抽出に注目が集まっている [1]．金融分野においても，決算短信・有価証券報告書・経済新聞記事・アナリストレポートなど投資家が入手可能なテキスト情報は常に溢れており，それら構造化されていないテキストデータから価値ある情報を自動で抽出することのニーズは大きい [2]．

金融テキストから情報を抽出し構造化する方法の 1 つとして，原因-結果の組から構成される因果関係を検出し，表現対として抽出することが考えられる．決算短信や経済新聞記事においては，因果関係を含む文が頻出する．例えば，「ウクライナ情勢をめぐる地政学的リスクの高まりで，エネルギーや原材料価格が上昇する」といったものだ．こうした因果関係を大量に収集することができれば，因果関係ネットワークの構築や因果系列の提示によって，投資判断の支援，ニュースイベントの波及効果の分析といった応用が可能となる [3, 4]．

金融テキストからの因果関係抽出においては，因果関係の存在を示す手がかりとなる表現の利用が有効だ [5, 6]．例えば，「レンタカー部門では，外出自粛の影響を受け，減収減益となりました.」という文の場合，手がかり表現「を受け,」を用いて，「外出自粛の影響」という原因表現と「(レンタカー部門では，) 減収減益となりました.」という結果表現を抽出できる．しかしながら，手がかり表現は因果関係の存在を示すとは限らない．「健康のため，運動する」という文の場合，「ため,」は原因結果関係の明示ではなく，「目的や期待の向かうところ」を表す．決算短信における具体例を表 1 に示す．

したがって，手がかり表現を含む文に対して，その文が因果関係を含むか否かを高い精度で判定することが望まれる．本研究では，金融ドメインの文書で追加で事前学習を行った金融 BERT モデルと，入力文の依存構造木に対して適用される Graph Attention Network (GAT) [7] を用いた判定モデルを提案する．決算短信と新聞記事データから構築したデータセットを用いて評価実験を行い，提案手法の有効性を確認した．

表 1　決算短信における，因果関係を含む文と含まない文の例．手がかり表現を太字で示す．

因果関係を含む文
・「ウクライナ情勢をめぐる地政学的リスクの高まり**で**、エネルギーや原材料価格が上昇するなど、国内経済の先行きは不透明な状況が続きました。」
・「レンタカー部門では、外出自粛の影響を**受け**、減収減益となりました。」

因果関係を含まない文
・「お客様のニーズに的確に応える**ため**、ビジネスパートナーとの連携強化を進めてまいりました。」

2 関連研究

テキストからの因果関係抽出に関する研究は数多く存在する．Girju [8] は英語の文書中の因果関係の存在を示す表現を調査し，それらを手がかりに自動で因果関係を検出・抽出する手法を提案している．坂地ら [9] は手がかり表現を含む候補文に対して，構文的な素性・意味的な素性を用いた機械学習手法 (SVM) でフィルタリングすることによって，因果関係を含む文を抽出している．

最近の研究では，BERT [10] のような事前学習済み言語モデルを利用するものが多い．The 4th Financial Narrative Processing Workshop の FinCausal-2020 Shared Task[1] では，Task1 が英文金融ニュースにおける因果関係を含む文の判定であった．様々な手法が提案されたが，その多くは BERT に基づくものである [11, 12, 13]．例えば Ionescu らは，BERT を含む 5 つの事前学習済みの Transformer ベースのモデルをアンサンブルする手法で全体 2 位のパフォーマンスを記録している．

深層学習に基づくモデルの学習では，一般に大規模な教師データセット作成のためのコストがかかるが，事前学習済み言語モデルの利点は，少数のデータでファインチューニングして精度の高いモデルを獲得できることである．一方で，関係抽出タスクにおいて，文の依存構造情報の利用が有用であることが知られている [14]．最近では，Graph Convolutional Network (GCN) [15] や GAT のようなグラフベースのモデルで依存木の構造を学習する手法が多く提案されている [16, 17]．本研究では，事前学習済み言語モデルとグラフベースのモデルの相補的な強みを活用することで，金融テキストから因果関係を抽出するための新しい因果関係文の判定モデルを提案する．

3 提案モデル

提案するモデルは，金融ドメインの文書で追加事前学習を行った金融 BERT モデルと，入力文のトークンの依存構造 (係り受けの構造) に対して適用される GAT により構成される．本節では，Graph Attention Network (GAT) の入出力について以下の表記を用いる．GAT において attention を実現する graph attention layer は，グラフ構造およびノード特徴集合 $\boldsymbol{h}^{(l)} = \{\boldsymbol{h}_1^{(l)}, \boldsymbol{h}_2^{(l)} ..., \boldsymbol{h}_{|V|}^{(l)}\}$, $\boldsymbol{h}_i^{(l)} \in \mathbb{R}^F$ を入力として，新たなノード特徴集合 $\boldsymbol{h}^{(l+1)} = \{\boldsymbol{h}_1^{(l+1)}, \boldsymbol{h}_2^{(l+1)} ..., \boldsymbol{h}_{|V|}^{(l+1)}\}$, $\boldsymbol{h}_i^{(l+1)} \in \mathbb{R}^{F'}$ を出力する．ここで $\boldsymbol{h}_i^{(l)}$ はノード i の l 層目の潜在表現を表す．また，$|V|$ はノード数，F, F' は隠れ層の次元である．

1) http://wp.lancs.ac.uk/cfie/fincausal2020/

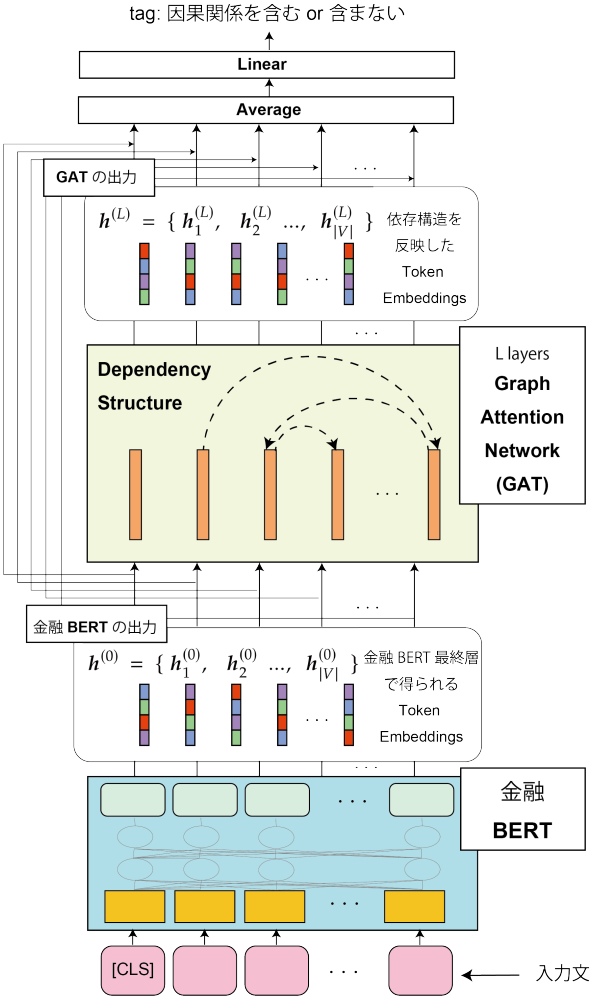

図 1 提案モデルの全体図．

提案モデルの概要は図 1 に示す通りである．まず入力文はトークン分割され，金融 BERT モデルに入力される．金融 BERT の最終層で得られるトークン毎の埋め込み表現は，GAT のノード特徴集合の初期値 $\boldsymbol{h}^{(0)}$ となる．各ノード (トークン) の特徴量は，GAT の各層で，依存構造木上の近傍となるノード N_i の特徴を用いて更新される．L 層目の出力のノード特徴集合 $\boldsymbol{h}^{(L)}$ は，skip 接続で $\boldsymbol{h}^{(0)}$ が加算された後に，全体の平均を取ることで読み出される．最後に，線形層を適用することによって，入力文が因果関係を含むか否かの予測を行う．

3.1 金融 BERT モデル

汎用言語コーパスで作成した事前学習モデルに，別のコーパスで事前学習を追加することを追加事前

学習と呼ぶ．対象タスクのドメイン周辺のコーパスで追加事前学習を行うことで，タスクでの精度向上が期待できる [18]．本研究では，日本語の汎用言語コーパスを用いて事前学習された BERT モデルに，金融分野のテキストを用いて追加事前学習を行うことで構築される金融 BERT モデルを用いる．

3.2 Graph Attention Network

GAT [7] の l 層目におけるノード i の近傍ノード N_i についてのメッセージ集約関数 $\boldsymbol{m}_{N_i}^{(l)}$ は，次式で表せる．

$$\boldsymbol{m}_{N_i}^{(l)} = \sum_{j \in N_i} \alpha_{ij} \boldsymbol{h}_j^{(l)} \tag{1}$$

ここで，attention スコアは以下で計算される．

$$\alpha_{ij} = \frac{e^{LeakyReLU(\boldsymbol{a}^T [W^{(l)} \boldsymbol{h}_i^{(l)} || W^{(l)} \boldsymbol{h}_j^{(l)}])}}{\sum_{k \in N_i} e^{LeakyReLU(\boldsymbol{a}^T [W^{(l)} \boldsymbol{h}_i^{(l)} || W^{(l)} \boldsymbol{h}_k^{(l)}])}} \tag{2}$$

$W^{(l)} \in \mathbb{R}^{F' \times F}$ は各ノードの重み行列，$\boldsymbol{a} \in \mathbb{R}^{2F'}$ は重みベクトルである．$\cdot^T$ は転置を，$||$ は連結を表す．ノード i の潜在表現は，次式のように更新される．

$$\boldsymbol{h}_i^{(l+1)} = \sigma(W^{(l)} \boldsymbol{m}_{N_i}^{(l)}) \tag{3}$$

ただし，σ は活性化関数である．

本研究では，文内のトークンをノード，2 つのトークン間の依存関係をエッジとするグラフ構造を GAT の入力とする．各ノード特徴は，金融 BERT モデルの出力として得られるトークン毎の埋め込み表現で初期化される．係り受け構造の解析は，日本語以外の言語への拡張性を考慮して，Universal Dependencies (UD) [19] に基づいて行う．

(式 3) のように，graph attention layer では距離 1 の近傍ノード N_i のみを考慮して潜在表現を更新する．したがって，L 層に重ねた GAT を適用することで，各トークンについて依存構造木上の距離 L 以内の近傍を明示的に考慮した潜在表現を得ることができる．このように構文情報を明示的に考慮し情報を集約する GAT を接続することによって，より精度高く因果関係の有無を判定できることが期待される．

4 実験

4.1 データセットと前処理

データセットを作成して評価実験を行った．用いたテキストデータは，決算短信・日本経済新聞記事の 2 種類である．決算短信データは，適時開示情報閲覧サービス[2]から PDF 形式で取得した．2012 年から 2022 年までに企業が発行した決算短信から無作為に抽出し，PDF 中の「経営成績に関する定性的情報」が記載された部分をテキストデータに変換して使用した．日経新聞記事データは 1995 年から 2005 年に発行された記事から無作為に抽出して使用した．各テキストデータは文単位で分割し，因果関係の存在を示す手がかり表現を含む文のみを使用した．本研究では，Sakaji らの研究 [6] で獲得された手がかり表現を用いた．それらを表 2 に示す．

表 2 因果関係の存在を示す手がかり表現．[6] に基づく．

を背景に　を背景に、　を受け、　を受けて
を受けて、　を受けております、　ため、　ためで、
ため」　ためで　ためだ。　ため。　ためであります。
に伴う　に伴い、　に伴い　で、　から、　により、
によって　により　による。　によります。
によっております。　によっています。　が響き、
が響いた。　が影響した。　が響く　が響いている
が響いている。　を反映して　を反映し、　このため、
このため　そのため、　そのため　その結果、
この結果、　をきっかけに　に支えられて

抽出された文に対してアノテーションを行った．決算短信データについては，1 人の評価者が各文に対して因果関係を含むか否かを示すタグを付与した．日経新聞記事データについては，5 人の評価者がタグを付与し，3 人以上が因果関係ありと判断したものを正例・そうでないものを負例とした．結果として，決算短信データでは 1958 件 (うち 1429 件で因果関係あり)，日経新聞記事データでは 2045 件 (うち 898 件で因果関係あり) の教師データを得た．

4.2 実験設定

金融 BERT モデルは，Suzuki ら [20] が公開する金融ドメインの文書で追加事前学習された BERT モデル[3] を利用した．このモデルでは，東北大学の乾研究室が公開する BERT モデル (Wikipedia 日本語記事を学習)[4] に対して，決算短信・有価証券報告書の 2 種類の金融コーパスで追加事前学習を行っている．

graph attention layer は 2 層に重ねて適用した．つまり，依存構造木上の距離 2 以内の近傍を明示的に考慮してノード特徴が更新される．また，学習の安定性のため 2 ヘッドの multi-head attention を用いた．

2) https://www.release.tdnet.info/
3) https://huggingface.co/izumi-lab/bert-base-japanese-fin-additional
4) https://github.com/cl-tohoku/bert-japanese

表 3　評価実験の結果．Precision, Recall, F1 の計算方法はマクロである．

	決算短信				日本経済新聞			
model	Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1
BERT	0.808	0.647	0.653	0.639	0.829	0.829	0.826	0.827
金融 BERT	0.866	0.850	0.781	0.805	0.879	0.879	0.878	0.878
BERT+GAT	0.852	0.767	0.744	0.748	0.889	0.890	0.886	0.888
金融 BERT ＋ GAT [ours]	**0.886**	**0.866**	**0.831**	**0.844**	**0.892**	**0.893**	**0.889**	**0.890**

入力文の依存構造の解析には，UD に基づいて設計された NLP フレームワークである spaCy[5] を用いた．2 つのトークン間における，主辞から修飾語への方向を持つ依存関係をエッジとするグラフ構造を GAT の入力とした．

提案モデルの有効性を示すために，以下の手法を比較手法として実験を行った．

BERT　Wikipedia の日本語記事を用いて学習された BERT モデル[4] の最終層の出力のうち，最初のトークン [CLS] に対応する出力に対して線形変換を適用することで予測を行う．

金融 BERT　金融 BERT モデル[3] の最終層の出力のうち，最初のトークン [CLS] に対応する出力に対して線形変換を適用することで予測を行う．

BERT+ GAT　提案モデルの金融 BERT[3] モジュールを BERT[4] に差し替えたモデルで予測を行う．

各データセットにおいて，64%を学習データに，16%を検証データに，20%をテストデータに割り当てた．学習データ内で 5fold の交差検証を行い，各 fold で検証スコアが最良の時のテストデータに対するスコアを記録し，結果はその平均値を用いた．なお，全ての手法において，BERT モデルは Transformer 層の上部一層のみをファインチューニングの対象とした．

4.3　結果と考察

実験結果を表 3 に示す．提案モデルである「金融 BERT+GAT」は，すべての指標・データセットで 3 つの比較手法の性能を上回った．また，「BERT」と「BERT+GAT」，「金融 BERT」と「金融 BERT+GAT」を比較すると，いずれのデータセットでも，すべての指標で GAT を用いる手法の方が精度が高かった．この結果は，GAT を用いることで明示的に構文情報を考慮する提案手法の有効性を示唆する．

決算短信データセットにおいて，提案手法を用いることによる，ベースライン手法 (「BERT」) からの精度向上が顕著であった．また，データセット間で「金融 BERT」と「BERT+GAT」の結果を比較すると，決算短信データセットでは，日経新聞記事の場合よりも，金融 BERT を利用することによる性能の向上が大きかった．このような結果の理由としては，金融 BERT の追加事前学習で用いる金融テキストコーパス中に決算短信が含まれていることや，決算短信における記述がニュース記事よりも専門度が高いものであるために，追加事前学習で金融特有の専門単語を反映することがとくに有効であったということが考えられる．全ての手法において日経新聞記事データの方が予測精度が高かったことからも，汎用言語コーパスで事前学習を行った BERT モデルの出力をそのまま利用する手法では，決算短信のような比較的専門性の高い文書中から因果関係を抽出することは難しいと示唆される．そのような場合でも，金融 BERT と GAT を用いる提案手法によって，抽出精度を大きく向上できることが確認された．

5　まとめ

本研究では，事前学習済み言語モデルとグラフベースのモデルの相補的な強みを活用することで，金融テキストから因果関係を抽出するための，新しい因果関係文判定モデルを構築した．金融分野のテキストから構築したデータセットを用いて評価実験を行うことで，提案手法の有効性が確認された．提案手法により，金融分野のテキスト中から，より精度高く因果関係を抽出することが可能となる．

今後の課題としては，英語や中国語において同様のデータセットを構築し，多言語に対応可能な判定モデルを構築することなどを考える．

5)　https://spacy.io/

謝辞

本研究はJSPS科研費JP21K12010，JST未来社会創造事業JPMJMI20B1，及びJSTさきがけJPMJPR2267の助成を受けたものです．

参考文献

[1] Jim Cowie and Wendy Lehnert. Information extraction. **Commun ACM**, Vol. 39, No. 1, pp. 80–91, 1996.

[2] B. Shravan Kumar and Vadlamani Ravi. A survey of the applications of text mining in financial domain. **Knowledge-Based Systems**, Vol. 114, pp. 128–147, 2016.

[3] Kiyoshi Izumi and Hiroki Sakaji. Economic causal-chain search using text mining technology. In **Proceedings of the 1st Workshop on Financial Technology and Natural Language Processing**, 2020.

[4] Kiyoshi Izumi, Shintaro Suda, and Hiroki Sakaji. Economic news impact analysis using causal-chain search from textual data. In **Proceedings of the AAAI-20 Workshop on Knowledge Discovery from Unstructured Data in Financial Services**, 2020.

[5] Christopher S.G. Khoo, Jaklin Kornfilt, Robert N. Oddy, and Sung H. Myaeng. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. **Literary and Linguistic Computing**, Vol. 13, No. 4, pp. 177–186, 1998.

[6] Hiroki Sakaji, Satoshi Sekine, and Shigeru Masuyama. Extracting causal knowledge using clue phrases and syntactic patterns. In **Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management**, 2008.

[7] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In **Proceedings of the 6th International Conference on Learning Representations**, 2018.

[8] Roxana Girju. Automatic detection of causal relations for question answering. In **Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering**, 2003.

[9] 坂地泰紀, 増山繁. 新聞記事からの因果関係を含む文の抽出手法. 電子情報通信学会論文誌, Vol. 94, No. 8, pp. 1496–1506, 2011.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, 2019.

[11] Sarthak Gupta. Finlp at fincausal 2020 task 1: Mixture of berts for causal sentence identification in financial texts. In **Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation**, 2020.

[12] Denis Gordeev, Adis Davletov, Alexey Rey, and Nikolay Arefiev. LIORI at the FinCausal 2020 shared task. In **Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation**, 2020.

[13] Marius Ionescu, Andrei-Marius Avram, George-Andrei Dima, Dumitru-Clementin Cercel, and Mihai Dascalu. UPB at FinCausal-2020, tasks 1 & 2: Causality analysis in financial documents using pretrained language models. In **Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation**, 2020.

[14] Aron Culotta and Jeffrey Sorensen. Dependency tree kernels for relation extraction. In **Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics**, 2004.

[15] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In **Proceedings of the 5th International Conference on Learning Representations**, 2017.

[16] Yuhao Zhang, Peng Qi, and Christopher D. Manning. Graph convolution over pruned dependency trees improves relation extraction. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, 2018.

[17] Jie Yang, Soyeon Caren Han, and Josiah Poon. A survey on extraction of causal relations from natural language text. **Knowledge and Information Systems**, Vol. 64, No. 5, pp. 1161–1186, 2021.

[18] Suchin Gururangan, Ana Marasovic, Swabha Swayamdi-pta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. Don't stop pretraining: Adapt language models to domains and tasks. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, 2020.

[19] Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. Universal Dependency annotation for multilingual parsing. In **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, 2013.

[20] Masahiro Suzuki, Hiroki Sakaji, Masanori Hirano, and Kiyoshi Izumi. Constructing and analyzing domain-specific language model for financial text mining. **Information Processing and Management**, Vol. 60, No. 2, p. 103194, 2023.