

論理的な課題を解くためのメモリ構造を導入した深層学習モデルに関する研究

新津 大我[†] 武田 敦志^{*}

本研究では、長期的な文脈情報を保持するメモリ構造と短期的な単語間の関係を捉える自己注意機構を統合した深層学習モデルを提案する。このモデルは、メモリ構造で文脈情報を蓄積・参照し、自己注意機構で関係性を解析することで、長期的な文脈保持と論理的推論を実現する。このモデルにより、従来のモデルが課題としていた長文や論理的な文脈を必要とする自然言語処理タスクにおいて、文脈保持と論理的推論の両面で性能を向上させることが可能となった。実験の結果、提案モデルは質問応答タスクにおいて従来モデルを上回る精度を達成し、自然言語処理分野における課題解決の新たな可能性を示した。

Deep Learning Models with Memory Architecture for Solving Logical Problems

Taiga Niitsu[†] and Atsushi Takeda^{*}

In this study, we propose a deep learning model that integrates a memory structure that retains long-term contextual information and a self-attention mechanism that captures short-term word-to-word relations. This model achieves long-term context retention and logical inference by storing and referencing contextual information in the memory structure and analyzing relationships in the self-attention mechanism. The proposed model improves the performance of both context preservation and logical inference in natural language processing tasks that require long sentences and logical contexts, which have been a challenge for conventional models. Experimental results showed that the proposed model achieved higher accuracy than the conventional model in the question-answering task, indicating new possibilities for solving problems in the field of natural language processing.

[†] 東北学院大学大学院人間情報学研究所, Graduate School of Human Informatics, Tohoku Gakuin University

^{*} 東北学院大学情報学部データサイエンス学科, Department of Data Science, Tohoku Gakuin University

1. 背景

近年、自然言語処理（NLP）は、ニューラルネットワーク技術の進化によって大きな発展を遂げている。リカレントニューラルネットワーク（RNN）や長短期記憶（LSTM）ネットワーク[1]は、逐次的な情報処理を得意とし、言語モデルや機械翻訳、感情分析など、多くのタスクで成果を上げてきた。また、Transformer[2]を基盤とするBERT[3]やGPT[4]などのモデルは、自己注意機構（Self-Attention）を採用し、文中の単語間の関係性を精密に捉える能力を持つ。この特徴により、Transformerは並列処理の効率性と高精度な文脈理解を両立し、自然言語処理の分野において重要な基盤技術となっている。しかしながら、これらのニューラルネットワークには、いくつかの課題が存在する。論理的な推論や長期的な文脈依存を伴うタスクでは、期待通りの性能を発揮できない場合がある。一般的なRNNやLSTMは、逐次的な情報を保持しながら処理する点で優れている一方、文脈内の情報を関連付けて統合する能力に改善の余地がある。一方、Transformerモデルは、自己注意機構によって文脈の関連性を精密に捉えることが可能であるが、長期的な情報を保持し続ける点でさらなる工夫が求められる。特に、長文や複雑な文脈を扱うタスクでは、それぞれのモデルが得意とする能力を超えた対応が必要となる。

本研究では、メモリ構造と注意機構を統合した深層学習モデルを提案する。この手法は、従来のRNNやLSTM、Transformerがそれぞれ抱える課題を克服することを目指して設計されている。具体的には、メモリ構造を用いることで長期的な文脈情報を動的に保持し、必要に応じて関連情報を適切に参照する。また、注意機構を活用することで短期的な単語間の関係性を高精度で解析し、複雑な文脈内の情報統合を実現する。このアプローチにより、長期的な依存関係や論理的推論が求められるタスクにおいても、一貫性のある処理が可能となる。

2. 関連研究

関連研究として、End-to-End Memory Networks[5]とDifferentiable Neural Computer（DNC）[6]が挙げられる。End-to-End Memory Networksは、外部メモリ[7]を利用して質問応答を行うニューラルネットワークである。主な特徴は、質問に関連する情報を外部メモリから段階的に参照し、回答を導く点である。この仕組みにより、複数の文やキーワードを基にした推論が可能となり、従来のRNNやLSTMが苦手としていた長期的な文脈を活用する能力が向上した。DNCはニューラルネットワークに外部メモリを組み合わせることで、柔軟な記憶操作を実現したモデルである。外部メモリを用いて情報を自由に読み書きできるため、大量の情報を効率的に処理し、順序情報、End-to-End Memory NetworksとDNCのいずれも、論理的推論を伴うタスクでは十分な精度を発揮できないという課題がある。一方で、Transformerは、その中心技術であ

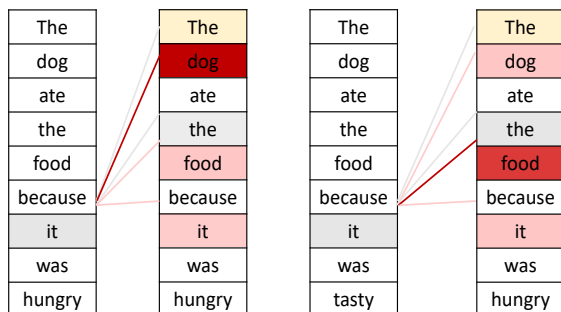


図 1 自己注意の可視化[8]

る自己注意機構によって、テキスト中の単語間の関係性を同時に評価する能力を持つ。この機構は、Query、Key、Value という 3 つのベクトルを使用して単語間の関連度を計算し、その結果得られる関連度スコアを基に、各単語が文脈内でどれだけ重要かを判断する。そして、重要度に応じて各単語の重みを動的に調整し、文脈全体の情報を統合する仕組みである。たとえば、「The dog ate the food because it was hungry」という文において、「it」が指す対象は文脈によって異なる。自己注意機構では、Query ベクトルを用いて「it」と他の単語 (Key ベクトル) の関連性を計算し、関連度スコアを得る。このスコアに基づき、Value ベクトルから関連情報を抽出する。図 1 に示されているように、「it」が「dog」を指している場合、その関連度スコアが高くなり、注意が「dog」に集中する。一方、別の文脈では「it」が「food」を指す可能性があり、このように自己注意機構は文脈依存的に関連性を動的に調整する能力を持つ。この特性により、Transformer は文脈全体を考慮しながら単語間の関係性を捉えることが可能となった。しかし、長期的な情報を保持し続ける能力にはさらなる改善の余地があり、長文の処理では性能が低下する場合がある。

3. メモリ構造を導入した深層学習モデルの提案

本研究では、メモリ構造と注意機構を統合した新しい深層学習モデルを提案しており、このモデルは長期的な文脈情報を蓄積・管理するメモリ構造と、短期的な単語間の関係性を動的に解析する自己注意機構を組み合わせることで、文中の単語間の関連性を精密に評価しつつ、過去の情報を保持して必要に応じて参照することにより文脈の一貫性を維持し、ストーリー全体の文脈構造を理解しながら質問に対する適切な回答を生成する能力を備えている。

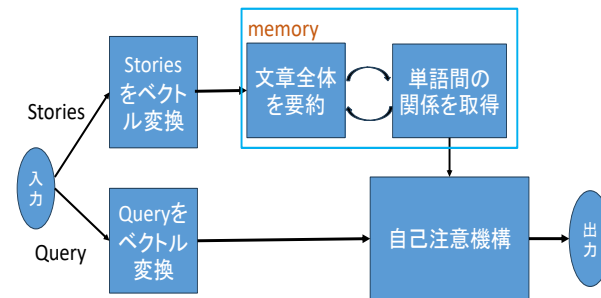


図 2 深層学習モデルの全体図

3.1 深層学習モデルの全体図

本節では、提案する深層学習モデルの全体構造を図 2 を基に説明する。本モデルは、入力データをストーリー (図 2 Stories) とクエリ (図 2 Query) の 2 種類に分け、それぞれをベクトル形式に変換した後、メモリ構造と自己注意機構を組み合わせる処理し、最終的な出力を生成する仕組みを持つ。まず、入力データはストーリーとクエリに分けられる。ストーリーは、文脈を形成する背景情報全体を指し、特定の質問に答えるための基盤となる。一方、クエリは、ストーリーに基づいて応答を生成するための質問を表す。これらのデータは、単語単位でベクトル形式に変換され、計算可能な数値データとしてモデルに入力される。さらに、入力データにはポジションエンコーディングが適用され、単語の順序情報が保持される。この際、クエリには単語の順序を明確に認識させるポジションエンコーディングが施され、ストーリーには単語の順序情報に加え、文の順序も認識できるように拡張されたエンコーディングが適用される。この処理により、ストーリーとクエリがそれぞれの構造を正確にモデルに伝える準備が整う。次に、ストーリーのベクトルはモデルの中核となるメモリ構造 (図 2 memory) に入力される。このメモリ構造は、ストーリー全体を要約しつつ、文脈情報を蓄積・保持する機能を持つ。また、文同士や単語間の関係性を動的に取得する仕組みを備えている。図 2 のメモリ構造部分が示すように、ストーリーが複数の文で構成される場合、これらの処理が繰り返し行われ各文の要約が蓄積され、それらの間の関連性が計算される。これにより、質問に答える際に必要な情報を効率的に取り出すことが可能となり、文脈の一貫性が維持される。さらに、メモリ構造によって保持された情報は、自己注意機構と連携する。この自己注意機構は、図 2 のクエリとメモリ構造からの矢印で示されるように、クエリを基にメモリ内に保存されたストーリー内の関連情報を動的に検索し、タスクに必要なデータを絞り込む役割を果たす。この過程では、クエ

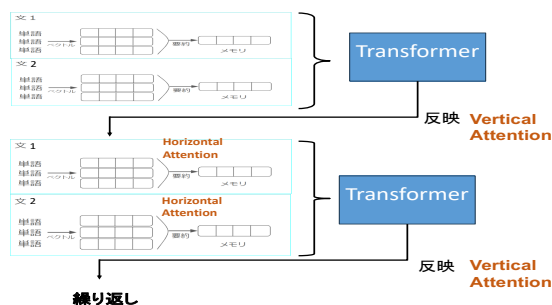


図 3 メモリ構造に情報の保持と更新

リとメモリ構造に蓄積された情報を比較し、それぞれの関連性が数値化される。関連性を示すスコアが高い情報ほど、出力に反映されやすくなるため、必要な情報が優先的に選択され統合される。この手法により、ストーリー全体の文脈とクエリの関連性を適切にモデル化することが可能となる。最後に、自己注意機構によって選択・精緻化された情報を基に最終的な出力が生成される。動的に抽出された関連情報が統合され、ストーリーとクエリの文脈的な関連性を反映した応答が生成される。この一連の流れにより、図2に示される深層学習モデルは、入力から出力までの全プロセスにおいて効率的に文脈情報を活用し、適切な結果を導き出す能力を実現している。

3.2 メモリ構造

本研究で提案する深層学習モデルにおけるメモリ構造は、ストーリーの文単位での情報統合と単語単位での詳細な情報処理を繰り返し適用することで、文内および文間の関係性を適切に解析し、長期的な文脈の一貫性を確保する仕組みを持つ。図3はこのメモリ構造の一連の流れを示している。まず、ストーリー内の各文は単語単位でベクトル形式に変換され、計算可能な数値データとしてモデルに入力される。次に、各文内の単語情報を集約し、文全体を要約する処理が行われる。この要約処理により、文内の単語間の情報が整理され、文単位の特徴ベクトルが生成される。生成された情報をメモリに保存する。この初期の各文の要約を基に、**Vertical Attention** が適用される。**Vertical Attention** は、文間の相互作用を解析し、文同士の因果関係や関連性を明確化する役割を果たす。この処理では、複数文間の情報が統合され、ストーリー全体の文脈が整理される。**Vertical Attention** の結果を基に、次に **Horizontal Attention** が適用される。**Horizontal Attention** は、文内の詳細な情報処理を行い、単語間の関係性を動的に解析する。この処理では、文中で重要な単語やその相互作用を評価し、文全体の情報

を精緻化する。その結果として得られた各文の要約がメモリに蓄積され、次の **Vertical Attention** に利用される。この蓄積された情報を基に、再び **Vertical Attention** が適用される。**Vertical Attention** は、メモリ内の既存情報と新たに追加された文の要約を統合することで、文間の相互作用をさらに深く解析し、ストーリー全体の文脈情報を更新する。この処理では、ストーリー全体を通じた因果関係や関連性が動的に整理される。**Horizontal Attention** と **Vertical Attention** の適用を繰り返すことで、文内と文間の情報が整理・統合され、メモリ構造内に効率的に蓄積されていく。この反復的な処理により、モデルは長期的な文脈情報を保持しつつ、詳細な単語間の関係性を把握することができ、この構造がストーリー全体の文脈の一貫性を確保しながら適切な応答を生成する基盤となる。

4. 実験・結果

4.1 実験

本研究では、提案モデルの性能を検証するために、質問応答タスク[9]を対象に実験を行った。このタスクは、ストーリーとクエリ（質問）が入力され、クエリに応じた正確な回答を生成する能力を評価するものである。このタスクを選定した理由は、提案モデルが長期的な文脈情報の保持と単語間の関係性解析を統合する設計であることに基づいている。具体的には、ストーリー全体を理解し、クエリに関連する情報を正確に特定する文脈の一貫性の保持と、クエリとストーリー内の単語間の関係性を捉えて適切な回答を導き出す能力という2つの観点から評価を行った。これらの特性は提案モデルの設計目的と直結しており、性能評価において重要な要素である。本実験では、質問応答タスクに特化したデータセットを用いて、提案モデルの性能を比較検証した。比較対象として採用したモデルは、Transformer、Memory Network、Differentiable Neural Computer (DNC) の3つである。それぞれのモデルは、文脈保持や単語間の関連性解析において異なる特徴を持つため、提案モデルの相対的な性能を評価するための適切な基準となる。評価指標として、生成された回答がクエリに対して正しいかを測定する回答精度 (Accuracy) を採用した。この指標を用いることで、提案モデルがクエリに対してどの程度正確な回答を生成できるかを定量的に評価した。

4.2 結果

表 1 は、タスク 1 からタスク 20 の結果についてその回答精度 (accuracy) を示している。実験の結果、提案モデルは一部のタスクで既存モデルを大きく上回る性能を示した。特に、タスク 1 (Single Supporting Fact) とタスク 4 (Two Argument Relations) では、それぞれ精度 1.000 という完璧なスコアを記録した。タスク 1 では、ストーリー

一内の単一の事実を正確に特定し、回答を生成する能力が実証された。また、タスク 4 では、ストーリー内の複数の文脈的関係を解析し、それらを統合して最適な回答を導き出す能力が確認された。さらに、タスク 9 (Simple Negation) では、否定表現の解釈が求められる中で、提案モデルは精度 0.922 を記録し、Memory Networks (0.885)、DNC (0.920)、Transformer (0.646) を上回った。この結果は、提案モデルが言葉のニュアンスを理解し、関連情報を正確に特定する能力を有していることを示している。また、タスク 16 (Basic Induction) では、未知の性質や属性を推論する能力を評価し、提案モデルは精度 0.486 を記録した。これは他のモデルを上回る結果であり、新たな文脈への適応能力が示された。一方、提案モデルはタスク 3 (Three Supporting Facts) とタスク 19 (Path Finding) において性能が低下する課題が見られた。タスク 3 では、ストーリーの文章量が多く、長期的文脈保持と複数の事実の統合が必要とされる中、提案モデルの精度は 0.210 にとどまった。タスク 19 では、ストーリー全体の文脈保持や複雑な論理的推論、経路の特定など複数のタスクが組み合わさる難易度の高い課題

表 1 各モデルにおける質問応答タスクの accuracy

accuracy				
task/model	transformer	memory network	DNC	proposal model
Single Supporting Fact	0.510	0.993	0.910	1.000
Two Supporting Facts	0.346	0.279	0.608	0.389
Three Supporting Facts	0.186	0.246	0.604	0.210
Two Argument Relations	0.767	0.957	0.990	1.000
Three Argument Relations	0.601	0.804	0.980	0.663
Yes/No Questions	0.794	0.911	0.930	0.606
Counting	0.696	0.826	0.900	0.851
Lists/Sets	0.716	0.821	0.940	0.908
Simple Negation	0.646	0.885	0.920	0.922
Indefinite Knowledge	0.721	0.414	0.900	0.851
Basic Coreference	0.801	0.907	0.960	0.918
Conjunction	0.759	0.996	0.950	0.772
Compound Coreference	0.943	0.942	0.960	0.944
Time Reasoning	0.625	0.972	0.890	0.920
Basic Deduction	0.584	1.000	0.720	0.511
Basic Induction	0.451	0.479	0.460	0.486
Positional Reasoning	0.493	0.571	0.670	0.524
Size Reasoning	0.685	0.736	0.950	0.898
Path Finding	0.112	0.113	0.350	0.126
Agents Motivations	0.976	0.808	1.000	0.917

であり、提案モデルは精度 0.126 を記録した。この結果は、これらのタスクが自然言語処理全体の課題であることを示している。以上の結果から、提案モデルは一部のタスクで既存モデルを上回る性能を示し、その設計の有効性が確認された一方で、長文文脈や複雑なタスクにおいてはさらなる性能向上が求められることが分かった。

5. まとめ

本研究では、長期的文脈保持と短期的関係性解析を統合した深層学習モデルを提案し、その性能を質問応答タスクを通じて検証した。提案モデルは、明示的な情報を特定するタスク 1 や、間接的な文脈を把握するタスク 4 において精度 1.000 を記録し、高い性能を示した。また、タスク 9 では、言葉のニュアンスを正確に理解し、それに基づく応答を生成する能力が評価され、既存モデルを上回る精度を達成した。さらに、タスク 16 では、未知の性質や属性を新しい文脈に適用する柔軟性が確認され、提案モデルの多様な推論能力を示した。一方で、長いストーリーを保持し、複数の事実を統合するタスク 3 や、論理的推論を伴う複数のタスクを統合したタスク 19 においては、さらなる性能向上が求められる結果となった。提案モデルは、長期的文脈を保持し関連情報を抽出する能力を有しているものの、タスク 3 のようにストーリーの文章量が多い場合には、メモリに保存すべき情報量がモデルの設計を超えてしまうことが課題として浮き彫りになった。また、タスク 19 では、ストーリー全体の文脈保持、複雑な推論、経路の特定といった複数の要素が組み合わさるタスクであり、すべてのモデルにとって困難であることが確認された。これらの結果から、タスク 3 とタスク 19 の課題は提案モデルに特有のものではなく、自然言語処理全体が直面している重要な研究課題であるといえる。本研究が提案する手法は、長期的文脈保持と短期的関係性解析を統合するアプローチとして、これらの課題解決に向けた新たな方向性を示している。特に、メモリ構造を活用した情報の保持・統合能力は、今後の自然言語処理において重要な役割を果たすと考えられる。この手法をさらに発展させることで、長文文脈の処理能力や多段階推論の強化が期待できる。本研究は、自然言語処理分野における長期的課題の解決に向けた一歩を示すものであり、今後のさらなる発展の基盤を提供した。

参考文献

- [1]S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, (1997).
- [2]Vaswani, A., et al., "Attention is all you need," Advances in Neural Information Processing Systems, (2017).
- [3]J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, (2018).
- [4]A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by

Generative Pre-Training," arXiv preprint arXiv:1810.04805, (2018).

[5]S. Sukhbaatar, J. Weston, and R. Fergus, "End-to-end memory networks," Advances in Neural Information Processing Systems, vol. 28, (2015).

[6]Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdomènech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, Demis Hassabis, "Hybrid computing using a neural network with dynamic external memory," Nature, Volume 538, Pages 471–476, (2016).

[7]J. Weston, S. Chopra, and A. Bordes, "Memory Networks," ICLR conference paper, (2015).

[8]H. Yun, T. Kang, and K. Jung, "Analyzing and Controlling Inter-Head Diversity in Multi-Head Attention," Applied Sciences, vol. 11, no. 4, pp. 1548, (2021).

[9]J. Weston, A. Bordes, S. Chopra, A. M. Rush, B. van Merriënboer, A. Joulin, and T. Mikolov, "Towards AI-complete question answering: A set of prerequisite toy tasks," arXiv preprint arXiv:1502.05698, (2015).