

二つの時系列データの関係を記述する 自然言語文生成手法の実測データ適用への取り組み

中野由加子 小林一郎

お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻

{g1920532, koba}@is.ocha.ac.jp

概要

本研究においては、人工データを用いて二つの時系列データの関係を説明する文生成を行う手法を提案し、実測された時系列データへの適用可能性について検証を行った。具体的には、初めに二つの時系列データの関係を捉えるため Transformer のクロス注意機構を拡張したモデルを用いて、人工的に作成した時系列データの挙動および関係性についての説明文生成の訓練を行う。そしてそのモデルを用いて、実測された時系列データの振る舞いについての説明文生成を行ない、人工データで学習したモデルの実測データへの転移学習の有効性について検証を行った。実測データについての説明文生成の精度は十分に高いものではなかったが、人工データの挙動に近い場合は正しく文生成を行うことが可能であることが確認された。

1 はじめに

Data-to-Text は、表 [1] や時系列チャート [2] など多様な表現形式での数値データにおける重要な部分について文生成を行う研究を指す。大量のデータから重要な部分を抽出して自然言語で説明することは人々のデータへの理解の助けとなるため、Data-to-Text の研究が注目されている。多くの Data-to-Text の研究は、数値データが観測されたドメインに特化した自然言語記述と同様な自然言語文を生成することを対象にしている [3], [4]。一方で、時系列データそのものが表現する意味や複数の時系列データ間の関係を捉えることも同じように重要にもかかわらず、そのような研究は少数であり、またそのうちのほとんどが単一の時系列データを説明対象とする [5]。単一の時系列データが与えられた場合、説明可能な情報はそのデータの挙動に限定される。しかし、同じ時間軸の複数の時系列データが与えら

表 1 動向の関係

協調	増加/増加, 減少/減少, 頂上/頂上, 凹み/凹み
反対	増加/減少, 減少/増加, 頂上/凹み, 凹み/頂上
片方平坦	増加/平坦, 減少/平坦, 頂上/平坦, 凹み/平坦 平坦/増加, 平坦/減少, 平坦/頂上, 平坦/凹み

れた場合、各データの個別の挙動に加え、これらのデータ間に生じる相互関係が分析対象として浮上する。相互関係の把握は、複数データ間での連動性や影響力を評価するための重要な要素となる。そのため、本研究では二つの時系列データを対象として説明文生成を行う。人工データセットを用いて説明文生成の手法を学習したモデルで、実測データについての説明文生成を行った。

2 データセット

本研究では、訓練のためのドメインフリーの人工データセットと評価のための実測された時系列データをもとにしたデータセットの 2 種類を構築した。

2.1 時系列データにおける説明対象

本研究で設定した説明対象の挙動を以下に示す。

動向/動向の関係 時系列データにおいて人が顕著に感じる動向として、「増加」、「減少」、「頂上」、「凹み」の 4 つが挙げられる (cf. [5])。二つの時系列データの動向の関係を上記の数値的变化量がある 4 種類の動向の組み合わせで構成する (表 1)。

時期 時系列データ長を 60 とし、0-19 を「初期」、20-39 を「中期」、40-59 を「末期」とする。

程度 二つの時系列データの動向の数値的变化量の比較結果を “more”, “less”, “considerable”, “slight” という単語を用いて説明をする。これらの単語を用いる場合の数値的条件を付録の表 6 に示す。

時間関係 時系列データが数値的变化量を取る動向を取り始める地点と取り終わる地点の比較結果を、“before”や“after”等の単語を用いて説明をする (付録の表 7)。

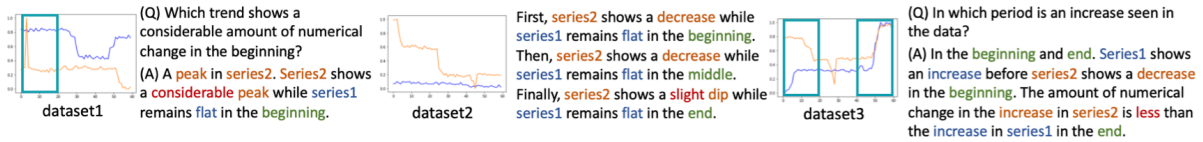


図1 人工データセット例

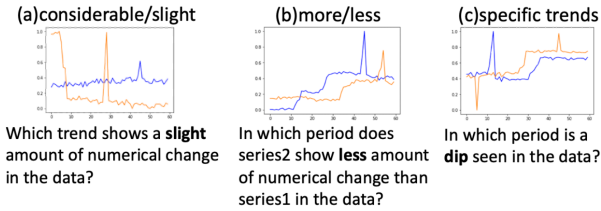


図2 質問文の種類とデータセット3の質問文例

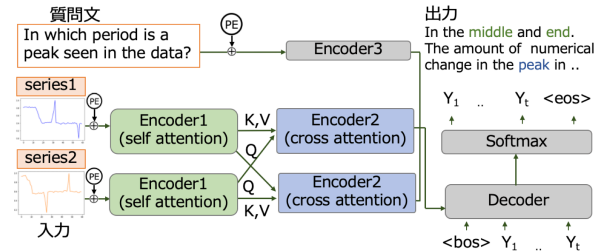


図3 提案モデルの概要

表2 人工データセット内訳

データセット	1,3	2
訓練	876,096	581,184
評価	5,184	5,184
テスト	3,456	3,456

表3 質問ごとのデータ内訳

	considerable	slight
気温	4	0
株価	5	4
コロナ	5	4

2.2 人工データセット

本研究では、説明対象とする挙動をとる時系列データを作成し、多様な分野に応用できるように正解文で特定の分野の専門用語の使用は避けることで、ドメインフリーの人工データセットを構築した。また、精度向上のために提案モデルの学習を3段階に分けて行い、それぞれの訓練のためのデータセットを作成した(図1)。データセットの時系列データは、初期, 中期, 末期において表1の内の1つの動向の関係をとる。「平坦」以外の動向をとる場合は、動向が明確に確認できるように0.2より数値的变化量が大きくなるようにした。また、時系列データを作成する際には、挙動が程度の数値的条件を満たす時期を決め、その時期を回答対象に必ず含めることで数値的变化量の比較結果を生成文に反映するようにした。決定した挙動を踏まえ、条件を満たすように乱数を発生させて時系列データを作成する。正解文の作成にはテンプレート(付録の図6)を用いている。正解文は、図1における動向や程度など色付きの単語を、質問文と時系列データの挙動を踏まえて置き換えることで作成する。質問は3種類作成し、(a) 数値的变化量が“considerable”または“slight”の条件を満たす動向を問う、(b) 2つの動向の数値的变化量の大小関係を問う、(c) 特定の動向について問うものとする(図2)。データセット内訳を表2に示す。

2.3 実測データセット

データセット内容 多様な実測データでの評価を行うために3つの異なる分野の実測データを用意した。1つ目はコロナ新規感染者数のデータで、厚生労働省が公開する全国47都道府県の新型コロナウイルス新規陽性者数の推移のデータのうち、2021年5月8日から2023年5月8日までににおける感染者数合計上位5県のデータを使用した。2つ目の気温データは、気候が異なる世界の5つの都市、温帯: Sydney (Australia), 冷帯: Anchorage (Alaska), 亜寒帯: Moscow (Russia), 乾燥帯: Riyadh (Saudi Arabia), 熱帯: Bangkok (Thai) の2022年1月1日から2023年12月31日の気温のデータを使用した。3つ目の株価データは、Yfinanceで取得したデータのうち、2024年9月20日段階で時価総額上位10社からランダム選んだ5社の2023年9月20日-2024年9月20日までの毎日の終値のデータを使用した。データには、butterworth filterを用いて高周波数成分を除去してノイズの影響を減少させる処理を行なった。またそれぞれの時系列データのとる値の大小関係により、どちらかのデータの挙動がほぼ平坦になるという状況を避けるために、スケーリングを行うことで両方の時系列データの挙動が観測できるようにした。

時系列データ抽出 学習に用いた人工データは、程度の数値的条件を満たす挙動を回答対象の時期において必ず1つ含むように作成した。そのため、実測データの評価データセットを作る際に、回答対象の時期において程度の数値的条件を満たす挙動を1つ含むデータを対象とした。対象のデータからそれぞれの分野のデータに対して3種類の質問ごとに

表4 人工データでの実験結果¹⁾

訓練	モデル	BLEU	ROUGE-L	METEOR	BERT	ppl.	T&P&D&C	T	P	D	C
3	提案モデル	0.606	0.730	0.826	0.962	1.73	0.419	0.665	0.720	0.741	0.681
	Transformer	0.346	0.628	0.563	0.927	2.20	0.00984	0.233	0.518	0.141	0.437
1&3	提案モデル	0.692	0.806	0.859	0.963	1.67	0.489	0.738	0.800	0.789	0.767
	Transformer	0.358	0.629	0.556	0.924	2.20	0.00984	0.261	0.529	0.151	0.438
2&3	提案モデル	0.706	0.788	0.861	0.965	1.69	0.345	0.764	0.792	0.849	0.765
	Transformer	0.371	0.568	0.633	0.932	2.12	0.0451	0.240	0.474	0.177	0.390
1&2&3	提案モデル	0.771	0.849	0.892	0.975	1.64	0.581	0.818	0.850	0.870	0.820
	Transformer	0.353	0.557	0.650	0.931	2.39	0.0388	0.231	0.460	0.173	0.391

ランダムに 10 ずつのデータを抽出するようにしたが、質問 (a) については “considerable と slight の数値的条件を満たす動向をとるデータが少数であったため、データ数が 10 未満となった (表 3)。正解文はデータセット 3 同様、テンプレートを用いて作成をした (付録の図 6)。

3 提案アーキテクチャ

提案モデルは、入力データの関係性を捉えるために Transformer [6] のクロス注意機構の拡張を行ない (図 3)、二つの時系列データそれぞれをエンコードするエンコーダが並列の構造をしている。Encoder1 はそれぞれの時系列データ自身を分析する。Encoder1 は transformer の self-attention メカニズムで生成されるベクトルである key, query, value は全て入力と同じ、つまり自分自身とする。一方で、Encoder2 は key と value は先ほどと同じであるが query については series1 と series2 を入れ替えることで、二つの時系列データの関係性を捉えられるようにしている。Encoder2 と質問文の解析を行う Encoder3 の出力を concat し、デコーダに入力する。デコーダは時系列データの解析結果と説明文の解析結果から説明文生成を行う。

4 実験

4.1 人工データでの訓練手法

人工データを用いた訓練において、時系列データの挙動についての質問回答文生成の精度向上のために訓練を 3 段階に分けて行い、それぞれの訓練に対してデータセットを構築した。以下にデータセットと訓練目的を示す。

訓練 1: 質問文の反映 データセット 1 は、時系列データ、特定の時期について問う質問文、正解回答文で構成する (図 1 左)。時期を指定した質問文への回答を通してデータの着目すべき視点を誘導して

質問文を反映した文生成の手法を学習する。

訓練 2: 単語ベクトルの獲得 データセット 2 は、時系列データと全ての挙動について説明をする正解文で構成する (図 1 中央)。データにおける全ての説明対象の挙動についての説明文生成を通して、データにおける挙動とそれを意味する単語の対応関係を学習する。

訓練 3: 質問回答文生成 データセット 3 は、時系列データ、挙動についての質問文、正解文で構成する (図 1 右)。訓練 1 と 2 を踏まえて、データの挙動についての質問回答文生成の手法を学習する。

4.2 評価指標

自動評価手法として BLEU, ROUGE-L, BertScore, METEOR, perplexity を使用する。正解文と生成文における 2 つの動向 (T), 時期 (P), 程度 (D), 時間的關係 (C) の正解率の計算を F1 スコアを用いて行なった。

4.3 訓練設定

エポック数を 200 とし、生成文を 1,000 イテレーションごとに評価データで評価し、評価値が最も高かった時のパラメータを用いてテストデータで評価を行う。訓練パラメータを付録の表 8 に示す。

4.4 結果

人工データを用いた訓練結果 3 段階の訓練を行うことで、精度が大きく向上した (表 4)。提案モデルの訓練 1&2&3 と train3 を比較すると、単語の正解率の 0.1-0.15 程度の向上が確認できる。また、どの学習方法においても Transformer が全体的に精度が低く、提案手法の精度を常に下回ることが確認できる。生成文が不正解であるもののうち、動向や程度などを正しく説明できていない例も見られたが、回答対象の挙動を取りこぼす場合や質問が反映できていない例も多く見られた。

1) X&Y は訓練 X の後に訓練 Y を行うことを意味する

表 5 実測データでの評価結果

データ	モデル	BLEU	ROUGE-L	METEOR	BERT ppl.	T&P&D&C	T	P	D	C	
気温	提案モデル	0.536	0.756	0.684	0.949	1.98	0.551	0.533	0.625	0.558	0.286
	Transformer	0.252	0.512	0.408	0.909	264	0.248	0.124	0.528	0	0.171
株価	提案モデル	0.619	0.746	0.722	0.955	1.87	0.654	0.674	0.761	0.542	0.607
	Transformer	0.314	0.620	0.627	0.938	2.26	0.370	0.348	0.562	0.00435	0.436
コロナ	提案モデル	0.594	0.832	0.713	0.959	1.74	0.631	0.634	0.698	0.566	0.531
	Transformer	0.345	0.535	0.518	0.926	2.28	0.310	0.317	0.512	0	0.280
全て	提案モデル	0.586	0.767	0.714	0.954	1.85	0.613	0.613	0.700	0.545	0.508
	Transformer	0.352	0.547	0.533	0.925	2.35	0.303	0.257	0.527	0.0161	0.345

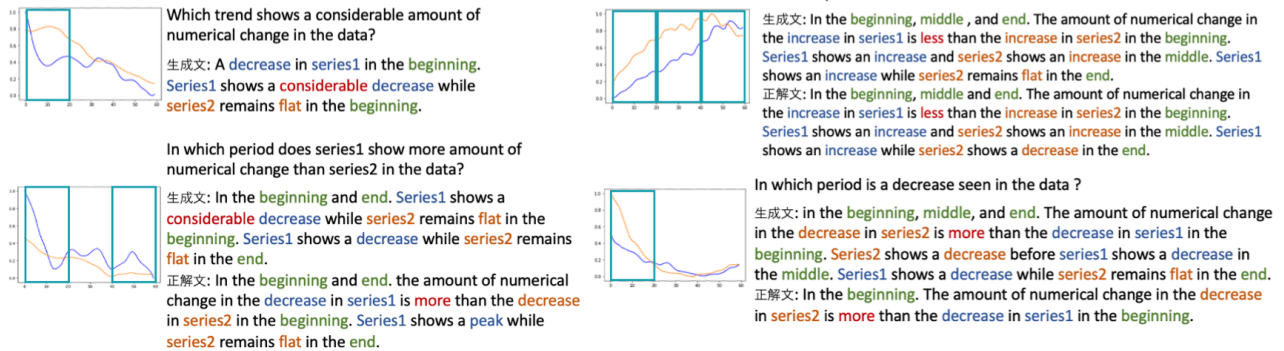


図 4 実測データでの生成文例

4.5 実測データの実験結果

結果を表 5 に示す。人工データでの結果 (表 4) から正解率は下がっているが、人工データを通して学習した挙動に近い場合は正しく回答文生成を行うことが確認できた。増加や減少については正しい説明文生成ができる場合が多く確認された。一方で、頂点と凹みは人工データでは鋭い挙動であったが本研究で用いた実測データにおいては緩やかな場合が多く、増加や減少と回答する場合が多かった。生成例 (図 4) の左上の場合、considerable な数値的変化量を取る動向を問う質問に対して、初期の series1 の減少を正しく回答できている。右上の場合、増加が見られる時期を問う質問に対して、全ての時期を回答できており、初期と中期の挙動を正しく説明できている。末期の series2 の減少を、平坦と回答したのは、series2 の数値的変化量があまり大きくないことが原因と考えられる。左下の場合、series1 が series2 より大きい数値的変化量を取る時期を問う質問に対して、初期と末期だと正しく回答をしている。しかし、series1 の減少の数値的変化量が大きいことから、series1 の “considerable” な減少と series2 の平坦の挙動だと誤って説明をした。右下は、減少が見られる時期を問う質問に対して、全ての時期を回答している。中期と末期においてはどちらのデータも数値的変化量が 0.2 未満で、平坦という判断に

なるが、中期においては目視でも series1 が微小に減少していることが確認できるため、この挙動を読み取って反映させたと考えられる。

本研究における実測データを用いた実験は、人工データでの制約により一部のデータに対してのみ説明文生成を行なった。また、ドメインフリーの人工データを作成したため実測データのような複雑な挙動ではなく単純な挙動をとるように時系列データを作成しており、一部の複雑な挙動について正しくない説明文生成を行なった。特定の分野で文生成を行う場合は、その分野のデータに似せて人工データを作成することや実測データを用いた fine-tuning を行うことで、より高い精度が期待される。

5 おわりに

本研究は、二つの時系列データの間を説明するために人工データセットを用いて学習をさせた提案モデルで実測データの挙動についての説明文生成を行なった。実測データに対する説明文生成を行なった結果の正解率は、十分に高いものではなかったが人工データで学習をした挙動に近いデータの場合は正しく回答ができていた。文生成を行う分野のデータの挙動に寄せた人工データセットの使用や実測データを用いた fine-tuning を行うことで、より高い精度が期待される。

謝辞

本研究は JSPS 科研費 23K28143 の助成を受けたものです。

参考文献

- [1] Deepanway Ghosal, Preksha Nema, and Aravindan Raghuveer. Retag: Reasoning aware table to analytic text generation, 2023.
- [2] Benny J. Tang, Angie Boggust, and Arvind Satyanarayan. Vistext: A benchmark for semantically rich chart captioning, 2023.
- [3] Jason Obeid and Enamul Hoque. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model, 2020.
- [4] Shankar Kantharaj, Rixie Tiffany Ko Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, and Shafiq Joty. Chart-to-text: A large-scale benchmark for chart summarization, 2022.
- [5] Harsh Jhamtani and Taylor Berg-Kirkpatrick. Truth-conditional captions for time series data. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 719–733, 2021.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

表 6 程度の単語の数値的条件

動向関係	程度	数値的条件
協調/反対	more/less	同時期の二つのデータの動向の数値的变化量の差が 0.15 より大きい
片方平坦	considerable	自身の数値的变化量に 0.6 をかけた値が、他の全ての動向の数値的变化量よりも大きい
片方平坦	slight	平坦な動向と自身以外の全ての動向の数値的变化量に 0.6 をかけた値が自身の動向の数値的变化量よりも小さい

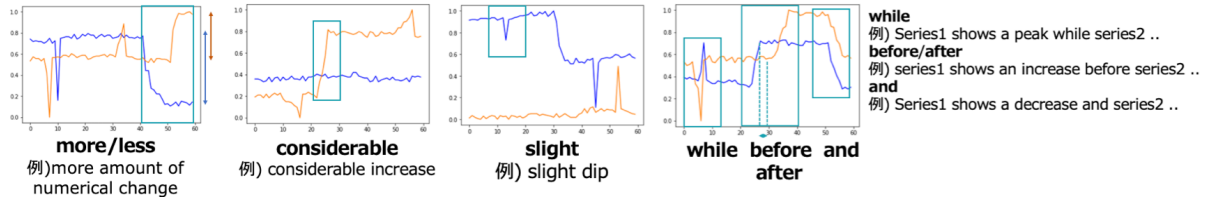


図 5 程度/接続詞の説明例

表 7 接続詞の条件

関係	end1 < start2	end2 < start1	左二つ以外
協調	before	after	and
反対	after	before	while
片方平坦	while	while	while

A. 程度に言及しないテンプレート

(協調/反対): series(X) shows (a/an) trend(X) (before/after/while/and) series(Y) shows a trend(Y) in the (period)
(片方平坦): series(X) shows (a/an) trend(X) while series(Y) remains flat in the (period)

B. 程度に言及するテンプレート

(協調/反対): the amount of numerical change in the trend(X) in series(X) is (more/less) than the trend(Y) in series(Y) in the (period)
(片方平坦): series(X) shows a (considerable/slight) trend(X) while series(Y) remains flat in the (period)

X=1, Y=2 or X=2, Y=1 trend(X) は series(X) の動向を示す

図 6 テンプレート例

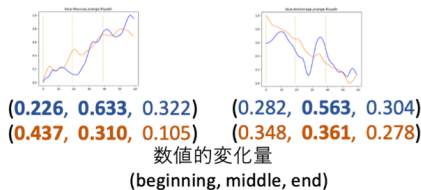
A 付録

A.1 人工データの正解文生成補足

本研究で設定した程度の単語の数値的条件を表接続詞の条件を表 7 に示し, seriesX が動向の数値的变化量を取り始める地点を startX, 取り終わる地点を endX とする (X=1 または 2). 図 5 の右の図の初期/中期/末期でそれぞれの時間関係の数値的条件を満たす場合を示す.

質問 (b) more/less の場合

(1) 数値的变化量の差から more/less の数値的条件を満たす時期を確認する



(2) (1) で確認した時期において数値的条件をみたま動向が一つのみかどうか確認をする

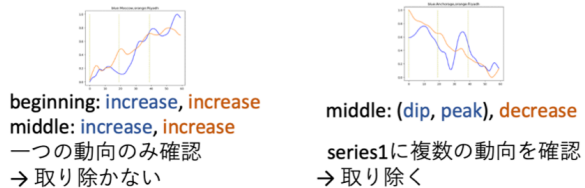


図 7 実人工データの抽出例

表 8 訓練設定

Embedding	128
隠れ層	512
損失関数	cross entropy
勾配法	Adam
学習率	0.0001
ドロップアウト	0.1
ウィンドウサイズ	8
バッチサイズ	8

A.2 実測データ評価データセット補足

時系列データは質問ごとに抽出を行なった. 質問が (b) の場合の時系列データの抽出例を図 7 に示す. 説明対象の挙動を含むデータのうち, 質問ごとに, 回答対象の時期の挙動に複数の動向が含まれる場合や学習していない動向の組み合わせを取る場合を取り除く. 図の (2) において左側の図は初期と中期の挙動が more/less の数値的条件を満たし, これらの時期においてはどちらも 2 つ以上の挙動をとらず, 増加のみを取る. 一方で, 右側のグラフは中期において more/less の条件を満たすが, 中期において凹みと頂点という二つの挙動を示す. そのため, 右側のデータは除去する.