

二つの時系列データの関係を記述する自然言語文生成への取り組み

中野由加子¹ 小林一郎¹

¹ お茶の水女子大学

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

概要

近年、表やグラフなど様々な数値データの表現形式を対象にした Data-to-Text の研究が盛んに行われている。多くの研究は対象ドメインにおけるイベントを説明するための自然言語文生成の中に数値データを取り込んだものとなっており、数値データそのものが表現する意味や複数の数値データ間の関係などを捉えるような自然言語文生成はほとんどない。本研究では、二つの時系列データの関係を捉え、記述する自然言語文生成手法の開発を目的とする。データセットとして、関連した動きをする部分がある二つの時系列データ、その関係を記述した自然言語文、カテゴリを作成した。作成したデータセットの訓練データで学習したモデルを用いて評価データについて文生成を行い、モデルの評価を行なった。

1 はじめに

近年、様々な分野における数値データの収集が容易になり、表や時系列チャートなど多様な表現形式での数値データについて自然言語で記述する Data-to-Text の研究が注目されている。多くの先行研究は、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。そのため、数値データ以外にもその数値が観測されたイベントを表す表題や発話者などのデータ発生の情報源といったコンテキスト情報を追加した自然言語文生成が提案されている [1]。

一方で、分析した数値解析結果を可視化ではなく言語化することによって、そのデータの内容を容易に把握することが可能になり、また対話システムや分析システムに組み込むことも可能になると考えられる。しかし、そのような数値解析に基づいた Data-to-Text はほとんど見受けられない。ドメインにこだわらず、時系列データの動きのみを説明する手

法として、Jhamtani ら [2] は、時系列データ中に特定のパターンが出現するかについての真理条件に基づいてキャプションを生成する手法（以下‘TRUCE’と呼ぶ）を提案している。TRUCE は、一つの時系列データにおいて特定のパターンが見つかった場合にそのパターンについて文を生成を行う。つまり、時系列データ特有のドメインを含んだ自然言語文生成を行うのではなく、時系列データそのものの動きを捉えた文生成を行っている。

本研究では、これに対して一つの時系列データに関する振る舞いを説明する自然言語文生成ではなく、二つの時系列データの関係を説明する文生成を対象とする。

2 複数時系列データの関係性記述

記述対象の関係性 時系列データにおいて人が顕著に感じる動きとして、「増加 (increase)」、「減少 (decrease)」、「頂上 (peak)」、「凹み (dip)」の4つが挙げられる (cf. [2])。二つの時系列データの振る舞いの組み合わせをこれらの4種類の動きで考えることも可能であるが、本研究では、二つの時系列データが共に増加、減少、頂上、凹みである場合の関係を捉え、それを自然言語で説明することを対象とする。また、時系列データについて説明する際には、「初期 (beginning)」、「中期 (middle)」、「末期 (end)」などの時期についても言及する傾向があるため (cf. [2])、対象とする4種類の動きに関する関係が見られる時期についても学習し、二つの時系列データの動きと発生時期について説明する文生成を行う。

単一の時系列データと複数の時系列データ説明文生成の相違点 複数の動きが見られる一つの時系列データの動きと時期について説明する場合には、どの動きに対する文生成を行なっても、その動きと時期を捉え、説明することが求められる。しかし、複

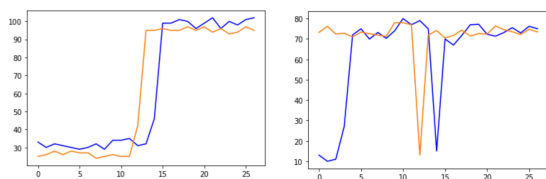


図1 データセット1

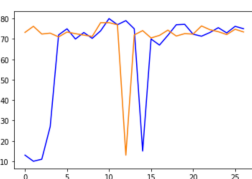


図2 データセット2

数の動きが見られる複数の時系列データの関係について説明する場合は、時系列データ間の関係性を捉えた上で、関係がある動きと時期についての説明を求められる。そのため本研究では、予め関係性のある二つの時系列データの動きを学習させることで、モデルが関係性のある動きを捉えやすくした。

3 データセット構築

二つの時系列データの関係性を説明するために、二つの時系列データの動きが共に増加、減少、頂上、凹みである場合の動きを学習させ、動きと時期を正しく捉えた文が生成できているかについて評価を行う。捉えたい二つの時系列データの動きをモデルに学習させるためのデータセット1と、学習させた動きを捉えることができているかの評価のためのデータセット2を作成した。その内容を以下に示す。

3.1 データセット1

捉えたい二つの時系列データの動きをモデルに学習させるために、入力として与える二つの時系列データ、訓練用の時系列データの正解カテゴリ、出力として期待される正解文のペアデータを96,000個作成した。その内、85%を訓練用データ、5%を評価用データ、10%をテスト用データとした。5,000step訓練を行なった内、1,000stepごとに評価データを用いて文生成を行い、BLUEスコアが最大の時のモデルを実験に使用した。モデルの学習用に訓練用・評価用データを使用し、モデルが訓練データを学習し、正しく文生成ができているかを評価するために、後述する実験1でテスト用データを使用した。

時系列データ 対象とする時系列データの関係となる4種類の動きとそれが生起する3種類の時期の12通りのデータを作成した。全体の時間幅を27と設定し、「初期」の場合は、0-8、「中期」の場合は9-17、「末期」の場合は18-26の範囲に二つの時系列データの関係を示す挙動が収まるようにしてデータを構築した。

選ばれた時期以外では、時系列データにほとんど変動がないようにした。動きが「増加」、時期が「中

表1 訓練設定

学習量	5000step
バッチサイズ	8
Embedding	64 次元
隠れ層	256 次元
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

期」である場合は、9-17の範囲内で増加傾向がある時系列データを作成する(図1参照)。作成したデータは、二つの時系列データの値が[0, 1]に収まるように正規化を行ってからモデルに入力する。

カテゴリ 二つの時系列データの関係性を捉えられるようにエンコーダを訓練するために、カテゴリを用意した。エンコーダの出力からカテゴリ判定を行う。上述した、動き4種類、時期3種類の組み合わせ12種類を0-11の数字で表したものをカテゴリラベルとして用いる。

正解文 動きと時期について言及する文を生成文の正解としている。データセットの作成を簡単にするために、”Both series1 and series2 (trend) in the (period).”というテンプレートを使用し、動き(trend)4種類と時期(period)3種類についての単語を置き換えることで正解文を生成した。

3.2 データセット2

3.1のデータセットで捉えたい動きと時期を学習したモデルを用いて、捉えたい動きの他にそれ以外の動きが含まれる時系列データについて説明文を生成する。評価用のデータセットのため、入力として時系列データ、出力として期待される正解文のペアを10,800生成し、全てを後述する実験2用のテストデータとした。

時系列データ 捉えたい動きを‘target trend’、捉えたい動きの時期を‘target period’とする。target trend以外の3種類の動きから1つを選択し、片方の時系列データにtarget period以外の時期のどちらかに選んだ動きを加え、合計288種類の二つの時系列データのペアを生成した。図2は、target trendが凹み、target periodが中期で、series1(青色の時系列データ)に増加の動きが初期に追加された例となっている。

正解文 捉えたい動きと時期のペア1つに対して、片方の時系列データにtarget trend以外の3種類の動きをtarget period以外の2種類から選んだ時期

表 2 実験結果

	BLEU	ppl	trend	period	trend period	category
実験 1	97.5	3.77	0.999	0.944	0.944	0.945
実験 2	62.4	3.14	0.703	0.545	0.393	0.571

に追加し、二つの時系列データを入れ替えたデータも作成した。この 12 種類について、同じ正解文を作成した。

4 提案モデル

4.1 概要

提案するモデルの概要を図 3 に示す。

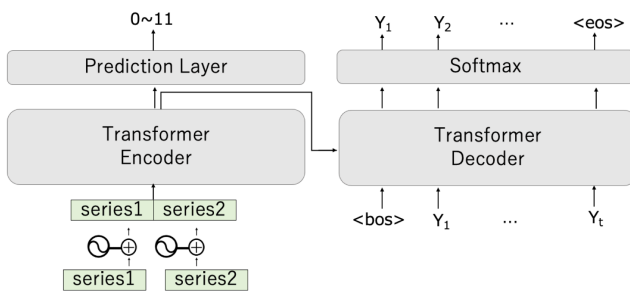


図 3 提案モデルの概要

モデルは Transformer [3] をベースに構築されている。入力二つの時系列データであり、Transformer を用いた多次元時系列データの異常検知を行う手法である TranAD [4] を参考に形式を決めている（4.2 節にて後述）。また、二つの動きの比較を可能にするために positional embedding をそれぞれに追加することにより時刻を共通化している。時系列データの特徴を上手くエンコーディングするために、先行研究 [1] のエンコーダの訓練手法を参考に、エンコーダからの出力を受け、上述した時系列データの動きと時期によって表される 12 種類のカテゴリを識別する層（図 3 中の Prediction Layer）を導入し、カテゴリの識別を訓練することにより、時系列データの関係を正しく捉えた埋め込み表現をデコーダに渡している。デコーダは Transformer のデコーダを採用している。エンコーダとデコーダをわけて訓練しており、エンコーダおよびデコーダの訓練時のハイパーパラメータの設定を表 1 に示す。

4.2 時系列データの入力形式

時系列データの振る舞いを捉えるために、「スライディングウィンドウ（sliding window）」と呼ばれる、ある程度の範囲を時間方向に少しずつシフトさせた特徴量によって時系列データを表現

表 3 実験 1 のカテゴリ・生成文の正解率

	カテゴリ				生成文			
	増加	減少	頂上	凹み	増加	減少	頂上	凹み
初期	1.00	1.00	0.948	0.904	1.00	1.00	0.791	0.815
中期	1.00	1.00	0.903	0.974	1.00	1.00	0.993	0.972
末期	1.00	1.00	0.940	0.700	1.00	1.00	0.862	0.876

することが多い。本研究においてもそのことを踏襲し、入力形式として時系列データにスライディングウィンドウを適用したものを採用する。ウィンドウサイズを K とし、時刻 t における値を $W_t = \{x_{t-K+1}, x_t, \dots, x_t\}$ とする。このとき、入力は、 w_t のリストとなり、長さ T の時系列データの場合、 $W = \{W_1, W_2, \dots, W_T\}$ と表される。この形式をとることで、変化量がわかりやすくなり、時系列データの動きが捉えやすくなる。

5 実験

実験 1 で、学習モデルが 3.1 の訓練用データで学習した通りに、二つの時系列データの動きを捉えて正しく文生成ができているか、データセット 1 のテストデータを用いて評価を行う。実験 2 では、捉えたい動きの他にそれ以外の動きが含まれる時系列データについて説明文を生成することで、より複雑な時系列データについても正しく文生成が行えるかを評価する。

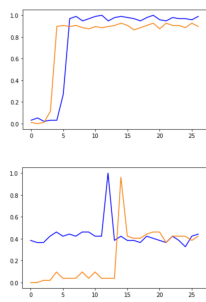
5.1 評価方法

生成文の質を評価するために、BLEU [5] と perplexity (ppl) を用いた。動き・時期を捉えられているかを確認するために、生成した文と正解文において、動き、時期、その両方を表す単語が一致しているかどうかを判定した。これらに加えて、Prediction Layer におけるカテゴリ判定と生成文の単語におけるより詳細な評価も行う。

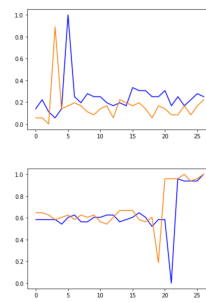
5.2 実験結果及び考察

実験 1 BLEU Score が非常に高くなっているが、今回はテンプレートを用いて正解文を生成しているため文同士に大差がなく、また動きと時期を 94 %以上の確率で正しく捉えられているためだと考えられる（表 2 参照）。

カテゴリ判定と生成文の正解率について 正しくカテゴリ判定ができたデータは全体の 94.5%となり、生成文において動きと時期の両方について正しく文生成ができていた確率 94.4%とほぼ一致した。



生成結果: Both series1 and series2
increase in the beginning.(増加・初期)
正解: Both series1 and series2
increase in the beginning.(増加・初期)
生成結果: Both series1 and series2
peak in the middle.(頂上・末期)
正解: Both series1 and series2
peak in the middle.(頂上・中期)



生成結果: Both series1 and series2
peak in the middle.(頂上・初期)
正解: Both series1 and series2
peak in the beginning.(頂上・初期)
生成結果: Both series1 and series2
dip in the middle.(凹み・中期)
正解: Both series1 and series2
dip in the end.(凹み・末期)

図4 実験1の生成文と(生成カテゴリ)/正解文と(正解カテゴリ)の例

表4 実験2のカテゴリ・生成文の正解率

	カテゴリ				生成文			
	増加	減少	頂上	凹み	増加	減少	頂上	凹み
初期	0.879	0.741	0.229	0.184	0.492	0.601	0.00118	0.014
中期	0.890	0.806	0.393	0.501	0.913	0.774	0.147	0.410
末期	0.904	0.924	0.300	0.217	0.527	0.502	0.021	0.290

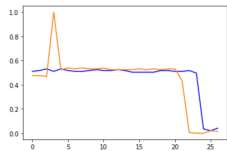


図5 動きが小さい例

表5 動きの大きさが異なるデータの割合・正解率

	same	half
データ割合	0.50	0.50
正解率	0.517	0.268

エンコーダの出力から文生成を行うため、カテゴリ判定の結果が生成される文における動きと時期の単語に直結すると考えた。しかし、頂上と凹みについては、カテゴリ判定と生成文の正解率が異なった(表3参照)。カテゴリ判定が正しくできていても、文生成が正しくできない場合(図4の右上)や、逆にカテゴリ判定が正しくできていなくても、文生成が正しくできている場合(図4の左下)が確認された。

実験2 実験1と比べて全体的に精度が下がっている(表2参照)。原因の1つとして、実験2におけるデータのうち50%が、捉えたい動きが小さくなったデータとなっていることが考えられる(表5参照)。実験2で用いたデータセットの時系列データは、加えた動きによっては正規化した際に、捉えたい動きを訓練時の半分ほどまで小さくしてしまう。例えば、図5においては、初期に頂上の動きを加えることで、末期の減少の動きの縦幅が全体の半分ほどまで小さくなっている。縦幅が小さくなっていないデータのうち51.7%は動きと位置について正しく文生成ができていたが、縦幅が小さくなっているデータについて正しく文生成が行えていたものは26.8%にとどまった(表5参照)。正規化をすることで、捉えたい動きの幅が小さくなってしまい、モデルが捉えたい動きだと判断することが難しくなった

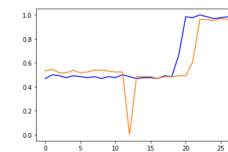


図6 時期だけでなく動きも正しく捉えられていない例

と考えられる。

カテゴリ判定と生成文の正解率について 正しくカテゴリ判定ができたデータは全体の57.1%となり、実験1に比べて大幅に下がった。カテゴリ判定において、時期だけでなく動きも正しく捉えられていないデータも見られた(図6)。これは、追加したデータの動きの影響を受けていることによると考えられる。また、全体的に生成文の正解率がカテゴリ判定の正解率からさらに下がっている(表4)。実験1よりも、エンコーダで動きが正しく捉えられていても文生成がうまくいっていないデータが多いとわかった。生成文と生成カテゴリ/正解文と正解カテゴリのペアをAppendixに記載した。

6 おわりに

本研究では二つの時系列データの関係性を捉えた文生成を行うモデルを提案し、作成したデータセットを用いてモデルの評価を行なった。実験1では、高い精度で二つの時系列データにおける、学習した動きを捉えた文生成を行うことができることが確認できた。実験2では、時系列データの動きがより複雑になり、動きの縦幅が学習時と異なる場合に動きを捉えることがより難しくなることが確認できた。正規化をすることで捉えたい動きが小さくなってしまいうため、捉えたい動きをうまく捉えて文生成ができるような工夫を考えたい。また、特に時期についての精度が下がってしまったので、訓練の手法等を検討して精度を高めていきたい。

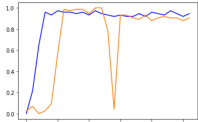
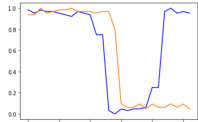
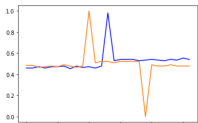
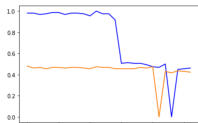
参考文献

- [1] Jason Obeid and Enamul Hoque. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model, 2020. <https://arxiv.org/abs/2010.09142>.
- [2] 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, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [4] Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. **Proc. VLDB Endow.**, Vol. 15, No. 6, p. 1201–1214, jun 2022.
- [5] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. In **Proceedings of the 40th Annual Meeting on Association for Computational Linguistics**, ACL ’02, p. 311–318, USA, 2002. Association for Computational Linguistics.

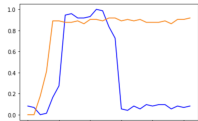
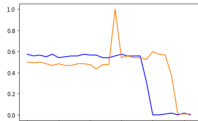
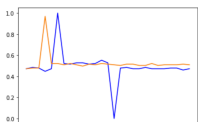
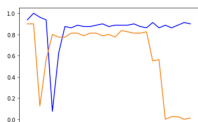
A 生成例

以下に実験 2 の生成例を示す。

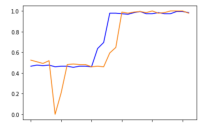
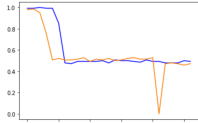
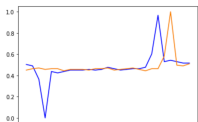
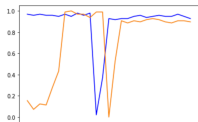
● カテゴリ判定も生成文も正解

	生成結果: Both series1 and series2 increase in the beginning. (増加・初期) 正解: Both series1 and series2 increase in the beginning. (増加・初期)		生成結果: Both series1 and series2 decrease in the middle. (減少・中期) 正解: Both series1 and series2 decrease in the middle. (減少・中期)
	生成結果: Both series1 and series2 peak in the middle. (頂上・中期) 正解: Both series1 and series2 peak in the middle. (頂上・中期)		生成結果: Both series1 and series2 dip in the end. (凹み・末期) 正解: Both series1 and series2 dip in the end. (凹み・末期)

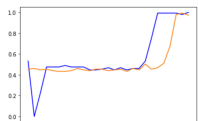
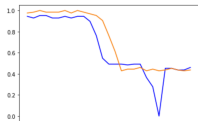
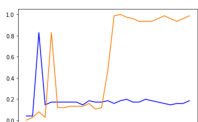
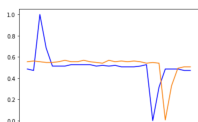
● カテゴリ判定は正解だが生成文は不正解

	生成結果: Both series1 and series2 peak in the beginning. (増加・初期) 正解: Both series1 and series2 increase in the beginning. (増加・初期)		生成結果: Both series1 and series2 peak in the middle. (減少・末期) 正解: Both series1 and series2 decrease in the end. (減少・末期)
	生成結果: Both series1 and series2 peak in the middle. (頂上・初期) 正解: Both series1 and series2 peak in the beginning. (頂上・初期)		生成結果: Both series1 and series2 dip in the end. (凹み・初期) 正解: Both series1 and series2 dip in the beginning. (凹み・初期)

● カテゴリ判定は不正解だが生成文は正解

	生成結果: Both series1 and series2 increase in the middle. (凹み・初期) 正解: Both series1 and series2 increase in the middle. (増加・中期)		生成結果: Both series1 and series2 decrease in the beginning. (凹み・中期) 正解: Both series1 and series2 decrease in the beginning. (減少・初期)
	生成結果: Both series1 and series2 peak in the end. (頂上・初期) 正解: Both series1 and series2 peak in the end. (頂上・末期)		生成結果: Both series1 and series2 dip in the middle. (凹み・初期) 正解: Both series1 and series2 dip in the middle. (凹み・中期)

● カテゴリ判定も生成文も不正解

	生成結果: Both series1 and series2 dip in the middle. (凹み・中期) 正解: Both series1 and series2 increase in the end. (増加・末期)		生成結果: Both series1 and series2 decrease in the end. (減少・末期) 正解: Both series1 and series2 decrease in the middle. (減少・中期)
	生成結果: Both series1 and series2 increase in the middle. (増加・中期) 正解: Both series1 and series2 peak in the beginning. (頂上・初期)		生成結果: Both series1 and series2 peak in the middle. (頂上・初期) 正解: Both series1 and series2 dip in the end. (凹み・末期)