

ChatGPT を用いた複数文章からの表生成

野田直哉¹ 村田真樹^{1,2}¹ 鳥取大学大学院 持続性社会創生科学研究科 工学専攻² 鳥取大学工学部附属クロス情報科学研究センター

m23j4039b@edu.tottori-u.ac.jp

murata@tottori-u.ac.jp

概要

本研究は、村田ら [1] の単語ベクトルとクラスタリングを用いた表生成と、ChatGPT を利用した表生成の精度を比較した。実験では、人手で作成した正解として定義した表との比較評価において、村田らの先行研究と ChatGPT を利用した手法の F 値が 0.61 と 0.64 であり、同程度の性能であった。また、正解のラベル (列の項目名) を用いて表を作成し場合の精度は F 値で 0.78 となった。

1 はじめに

近年、Web 上のテキストデータの構造化や、情報抽出に関する様々な手法 [2][3] が提案されている。その一環として、本研究では関連した文章から項目ごとに情報を整理することで表の作成を行う。一般的に要約は、長文から情報の取捨選択をし、それ以外の情報を捨てることになる。しかし、捨てられた情報の中にも場合によっては必要な情報となる場合もある。そこで、情報の取捨選択ではなく項目ごとに整理することで、もとの文章の情報量を維持しつつ、読み手は項目に注目することで素早く必要な情報を取得できる。そこで図 1 の文章を、表 1 のように重要情報を自動で抽出し表の形で整理することで、人手で文章を読む労力を減らし、可読性と利便性の向上をはかる。

村田ら [1] の研究では表データの学習なしで文章をベクトル化しクラスタリングを行うことで表を作成していた。また、類似の研究として、赤野ら [4] の研究では単語を抜き出し、単語単位での表生成を行っているほか、村田ら [5][6] の研究では、ChatGPT と表生成技術を用いて、株価記事に対する分析を行っている。本研究は、この表生成の手法を ChatGPT に置き換え、既存の手法との精度を比較する。

本研究の主な主張は以下の 2 点である。一つ目に、

この実験で用意した基準となる表との比較評価において、先行研究と ChatGPT を利用した手法の F 値が 0.61 と 0.64 であり、同程度の性能であった。二つ目に、文の分類精度のみに着目し、正解のラベルで表を作成し場合の精度は F 値で 0.78 となった。

文章1	文章2	文章3
身長183cm,... 出身は... 最高位は...	身長175cm,... 血液型は... 最高位は...	現役時代の,... 血液型は... 最高位は...

図 1 入力例 (力士)

表 1 出力例 (力士)

	身長	最高位
文章 1	身長 183cm, 体重 191kg	最高位は東十両 12 枚目
文章 2	身長 175cm, 体重 130kg, 血液型は O 型	最高位は西幕下 2 枚目
文章 3	現役時代の体格は身長 179cm、体重 149kg, 血液型は AB 型	最高位は東幕下 4 枚目

2 単語ベクトルを用いた手法

村田らの研究では、複数の文章から重要情報を抽出し表の形で整理して表示していた。図 2 と図 3 にあるように、文の単語のベクトルの平均を文のベクトルとして、クラスタリングし、各クラスタの中から重要項目となりうるクラスタを選別し表の形で整理していた。

- 事前準備として Wikipedia から FastText[7] を用いて単語ベクトルのデータベースを作成する。
- 複数文書を句点区切りの文に分割する。
- 文を形態素解析により名詞のみを抽出する。
- 名詞の単語ベクトルを平均し文ベクトルを求める。
- 文ベクトルを階層化クラスタリングでクラスタリングする。

表2 ラベル生成のプロンプト

対象の記事を示す	
入力	<記事>
すべての文にラベル付けできる十分な数のラベルを生成	
入力	表を作成する際のラベルを 30 個程度生成出力してください
出力	<出力されたラベル 30 個>
一度厳選を行い、ラベルの数を調整	
入力	出力されたラベルの厳選をお願いします。
出力	<厳選されたラベル>

- クラスタリング結果からセルが最も多く埋まっているクラスタ数のクラスタを採用する。
- クラスタリングの結果を表にまとめる。

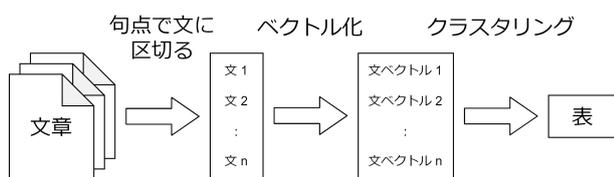


図2 単語ベクトルを用いた手法の手法

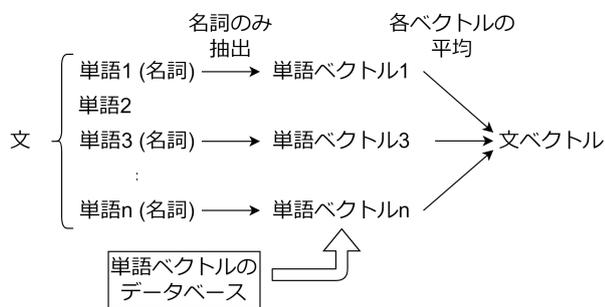


図3 単語ベクトルを用いた手法 文ベクトルの作成

3 提案手法

本研究では、ChatGPT を用いて表生成を行い、手順としては表に用いるラベルを生成と、そのラベルで、各文へラベル付けをしている。

表2 はラベルを生成させるプロンプトで、すべての文にラベル付けできる十分な数のラベルの生成を指示している。しかし、ラベル数が多い場合、各文でつけられるラベルにまとまりがなくなる。そこで、その中から必要度が高いと考えられるラベルを適宜、厳選しラベル付けで使用できるように出力させている。

表3 は文にラベル付けをするプロンプトで、厳選

表3 各文へのラベル付け

<厳選されたラベル>でラベル付けを行う	
入力	<厳選されたラベル> を用いて以下の文のラベリングをお願いします。 “xml < bun labels=”?”> 文 1 </bun > < bun labels=”?”> 文 2 </bun > < bun labels=”?”> 文 3 </bun > “
出力	“xml < bun labels=”ラベル a”> 文 1 </bun > < bun labels=”ラベル b”> 文 2 </bun > < bun labels=””> 文 3 </bun > “

ラベル付けされなかった文を使用していないラベルでラベル付け

入力	<出力されたラベル 30 個> を用いて以下の文のラベリングをお願いします。 “xml < bun labels=”?”> 文 3 </bun > “
出力	“xml < bun labels=”ラベル c”> 文 3 </bun > “

されたラベルを用いて、分割した文にラベル付けを指示している。このとき、どのラベルとも関連がなくラベル付けされない文が出現することがあり、これらの文は、使用されなかったラベルを用いて再度ラベル付けを行っている。

4 実験

4.1 実験方法

ChatGPT の対話システムを用いて、表生成を行う。使用するモデルは GPT4 である。実験に用いるデータは web から収集した日本語記事を種類ごとにまとめ記事群とし、15 個の記事群を用いる。また各記事群に約 20 記事が含まれ、一つの記事群から表を一つ作成する。これら各記事の長さはおおよそ、370 字、8 文程度となっている。

これらの記事の全ての文には、正解として表のどこに割り当てられるか、情報をラベルとして持ち、このラベルはあらかじめ人手で決めている。表4 は実

際に使用する記事中の文と、そのラベルである。

表4 エアコン記事の文とラベル

発売日
11月1日より発売する。
3月下旬より発売する。
10月下旬より順次発売する。
掃除・手入れ
さらに、吹き出し口周辺のルーバーやダストボックスなどは簡単に取り外し、水洗いが可能。
前面パネルにはホコリが付きにくい帯電防止加工を施し、清潔性を向上させている。
20畳相当（4.0kW以上/200Vタイプの場合）の高い集じん能力を実現した。

4.2 評価方法

実験によって得られた表を、正解と定義した表と比較し評価する。この評価に基づいて、再現率 (Recall)、適合率 (Precision)、F 値を算出することで情報抽出の精度を求める。生成された表の精度を以下の手順で算出する。この実験では表で使用されるラベルが固定されていないため、正解と出力された表の列同士をマッチングする必要がある。ラベル名でのマッチングは表現の揺れを考慮すると、正常に判定することができないため、含まれている文が最も近い列同士をマッチングさせている。

1. 図4において、正解のデータの列1に注目する。
2. 列1と出力の列a~xのそれぞれを比較し、式3をもとにF値を求める。ただし、Aを正解の列に含まれる文の集合、Rを結果の列に含まれる文の集合とし、これらの絶対値は集合に含まれる要素数とする。
3. 出力の列a~xでF値が最も高いものを列1とマッチングし、評価する。
4. 上記を正解のデータの他の列についても行う。

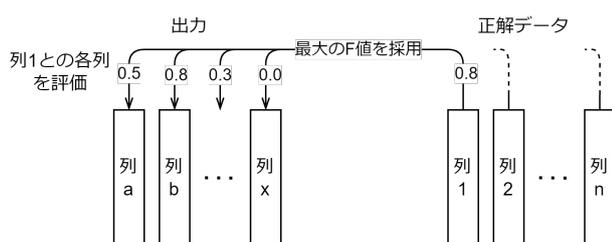


図4 評価方法

$$\text{Recall} = \frac{|A \cap R|}{|A|} \quad (1)$$

$$\text{Precision} = \frac{|A \cap R|}{|R|} \quad (2)$$

$$F = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} \quad (3)$$

5 実験結果

本研究では、単語ベクトルを用いた手法と提案手法において、15個の表を作成しそれぞれ評価した。表5に、出力された各表の結果の平均を示す。先行研究と ChatGPT を利用した手法の F 値が 0.61 と 0.64 であり、同程度の性能であった。

表5 実験結果に対する評価

	単語ベクトルを用いた手法	提案手法
適合率	0.61	0.55
再現率	0.75	0.81
F 値	0.61	0.64

5.1 ラベル生成の例

実験で作成されたラベルの例を表6に示す。正解の表でのラベルと出力されたラベルでは「価格」、「発売日」、「省エネ・エネルギー効率/消費電力」など共通の事柄が出力されている一方、「掃除・手入れ」、「AI」、「掃除・手入れ」など共通していないラベルが含まれている。

表6 生成されたラベルの例 (エアコン記事)

正解の表でのラベル	出力されたラベル
価格	価格
発売日	発売日
省エネ	エネルギー効率/消費電力
ラインアップ	モデル名
カラー	色/デザイン
冷暖房性能	冷房能力 (kW)
気流の制御	暖房能力 (kW)
サイズ	適用畳数 (冷房時)
掃除・手入れ	適用畳数 (暖房時)
除菌・清浄	空気清浄機能
除湿	主な特徴
デザイン	
AI	
人物検知	
メーカー発表	
他の機器との連携	

5.2 ラベリング結果の例

実験で得られたラベリングの結果を表7に示す。ラベル「発売日」に関しては表の例では正解と一致しているが、出力のラベル「色/デザイン」で判定された文が正解の表では「掃除・手入れ」「カラー」「デザイン」の3つのラベルで分類されている。このように、文へのラベル付けは、用いるラベルに依存することがわかる。

表7 ラベリングされた文の例

正解/出力	発売日
発売日	11月1日より発売する。
発売日	3月下旬より発売する。
発売日	10月下旬より順次発売する。
正解/出力	色/デザイン
掃除・手入れ	さらに、吹き出し口周辺のルーバーやダストボックスなどは簡単に取り外し、水洗いが可能。
カラー	ボディカラーは、パウダースノウとボルドーレッドの2色を用意した。
デザイン	風の流れをイメージして曲面を基調とした「ウェーブデザイン」の前面パネルを採用。

6 考察

6.1 生成されるラベルへの依存

5節では、GPT4を用いて、ラベル生成と文へのラベル付けにより、表生成をした。実験では表7のように、文へのラベル付けは、用いるラベルに依存することがわかる。同一の文書であっても、用いるラベルにより多様な表が生成できることにより、一概に正解を定義することは困難で、表生成の性能測定は難しいものと考えられる。

6.2 ラベル付けの性能

実験で、F値は正解とした表と出力された表のラベルの差に依存することを確かめた。文の分類精度のみに着目し、ラベル付けの精度のみを抽出するため、正解のラベルを用いて、GPT4でラベル付けのみを行った。このとき評価は表8の「ラベル付けのみ」である。このことより、ラベル生成の問題が解決できた場合の精度はF値で0.78程度出せると考えられる。

表8 ラベル付け性能の評価

	単語ベクトル	ラベル生成+ラベル付け	ラベル付けのみ
適合率	0.61	0.55	0.70
再現率	0.75	0.81	0.88
F値	0.61	0.64	0.78

7 おわりに

近年、Web上のテキストデータの要約や、情報抽出に関する様々な手法が提案されている。その一環として、本研究では関連した文章から項目ごとに情報を整理することで表の作成を行った。一般的に要約は、長文から情報の取捨選択をし、それ以外の情報を捨てることになる。しかし、捨てられた情報の中にも場合によっては必要な情報となる場合もある。そこで、情報の取捨選択ではなく項目ごとに整理することで、もとの文章の情報量を維持しつつ、読み手は項目に注目することで素早く必要な情報を取得できる。

本研究は、村田ら[1]の単語ベクトルとクラスタリングを用いた表生成と、ChatGPTを利用した表生成の精度を比較した。実験では、人手で作成した正解として定義した表との比較評価において、村田らの先行研究とChatGPTを利用した手法のF値が0.61と0.64であり、同程度の性能であった。また、正解のラベル(列の項目名)を用いて表を作成した場合の精度はF値で0.78となった。

参考文献

- [1] Masaki Murata, Kensuke Okazaki, and Qing Ma. Improved method for organizing information contained in multiple documents into a table. 自然言語処理, Vol. 28, No. 3, pp. 802–823, 2021.
- [2] Chia-Hui Chang, Kaye Mohammed, Girgis Moheb, Ramzy, and F. Shaalan Khaled. A survey of web information extraction systems. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, p. 1411–1428, 2006.
- [3] Ali Farman, Kwak Daehan, Khan Pervez, Shaker Hassan A. Ei-Sappagh, Islam S. M. Riazul, Park Daeyoung, and Kwak Kyung. Sup. Merged ontology and svm-based information extraction and recommendation system for social robots. *IEEE Access*, Vol. 5, No. 7962152, 2017.
- [4] Akano Hokuto, Murata Masaki, and Ma Qing. Detection of inadequate descriptions in wikipedia using information extraction based on word clustering. *IFSA-SCIS 2017*, pp. 1–6, 2017.
- [5] Masaki Murata. Content analysis of items in newspaper data using table arrangement technology and chatgpt for stock price prediction. In *Proceedings of The 22nd International Conference on Information & Knowledge Engineering on CSCE 2023*, pp. 1–8, 2023.
- [6] Masaki Murata. Leveraging chatgpt and table arrangement techniques in advanced newspaper content analysis for stock insights. *Big Data, Data Mining and Data Science: Algorithms, Infrastructures, Management and Security*, 2024. (submitted).
- [7] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword in-

formation. **arXiv preprint arXiv:1607.04606**, 2016.