ChatGPT を用いた重要語句の思考連鎖による複数文書要約

小西 健太*1
指導教員：長谷川 卓也*2
Email: hasegawa@tachibana-hs.jp

*1：京都橘高等学校３年
*2：京都橘高等学校

◎Key Words　　複数文書要約，ChatGPT，プロンプト，思考連鎖

1.　はじめに

ある語句をインターネットで検索すると，数多くの記事がヒットする。一つの記事のみを参考にすると，理解に偏りが生じてしまうため，複数の記事を参考にすることが求められる。語句によっては多面的な意味があり，着眼点の異なる説明を行う記事が複数存在する場合がある。それらを一つの文書に要約し，語句に関して総合的に理解したいときには「複数文書要約」の技術が役に立つ。複数文書要約の技術は自然言語処理の研究の一分野として注目されている(1)。要約においては冗長性を排除し，必要な情報を精度高く合成することが求められる。

ChatGPT(2)をはじめとする大規模言語モデルの進歩は革新的であり，多くの人々の注目を集めている（図１）。特にプロンプト[1]の推敲が生成に大きく影響する。目的に応じて新しいプロンプトを学び，研究することは重要である。プロンプトを工夫することで冗長性を減らし，精確な要約が可能になると考えた。また人工知能を活用する際は，生成の安定性も要約において大切な観点となる。

先行研究として，Li らによる“Guiding Large Language Models via Directional Stimulus Prompting”(3)　という論文がある。この論文では，教師あり Fine-tuning された LM[2]を利用して単一文書から重要語句を抜き出し，より良い文書要約を出力するという手法が提案されている。しかし，このような作業では高度な技術が求められるため，一般の人々への普及は困難であると考えた。そこで，一般の人々にとっても馴染み深く，手軽に使うことができる ChatGPT を利用して重要語句の抽出をし，抽出した語句について説明を行わせることで，同様の効果を得られると考えた。そして，Wei らによる“Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”(4)　という論文では，段階的な推論ステップを示すことで，モデルは複雑な推論を行いやすくなることが実証されている。重要語句について説明を行わせ，要約を行わせるプロンプトにすることで，連鎖的に思考を行う CoT[3]（Chain-of-Thought）プロンプティングの考え方を，本研究のプロンプトに取り入れることができると考えた。複数文書要約において Bao らは“Chain-of-event prompting for multi-document summarization by large language models”(5)という論文を発表している。CoT プロンプティングを応用し，各記事から出来事となる文を一つずつ抽出・一般化し，要約を行なっている。本研究では，一つの語句について解説を行う記事の要約を目標としているため，複数記事から重要語句を抽出するプロンプトを設計した。

手軽に使うことができる ChatGPT を用いて，一段階目は複数記事から一括で重要語句を数個抽出する。二段階目では抽出した語句を説明させ，複数文書の要約を行う一連の手法を新たに考案した。これを“Chain-of-Words-Thought prompting”（以後，CoWT プロンプティングと呼ぶ）と名付けた。

よって，CoWT プロンプティングがより簡潔かつ精確で安定性のある複数文書要約であるかどうかを研究した。

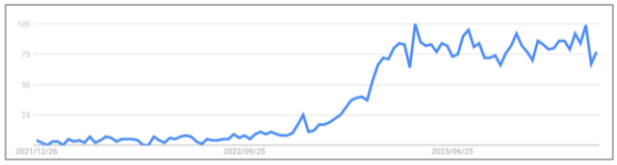

図１　「大規模言語モデル」検索の人気度推移（2022/01/01 - 2024/01/01）
出典：Google トレンド「Large Language Model」

2.　研究方法

2.1　使用した題材

本研究で行う手法において，最も効果的な題材は専門用語であると考えた。理由は，専門的な題材では知識が深く広範であり，異なる視点やアプローチが存在し得るためである。記事の多面性を活かし，記事を要約させるプロンプトに変化を加えることで，要約の生成の精度も変化すると考えた。この考えは，要約する際に記事の文書・表現の取捨選択を ChatGPT がプロンプトを元に行っている性質に由来する。したがって，本研究では，多面的な説明があり得る語として，「コンピュータの CPU」を選んだ。CPU は高等学校情報科「情報Ⅰ」において登場する語である。しかし，中学生・高校生にとって馴染がない。事前のアンケートでは，CPU の説明がある程度できる割合は約一割という低い結果が出ており，CPU の理解度は低い[4]。教科書では，主にコンピュータの基本構成の一部である CPU として説明がなされている。一方で，インターネット上では CPU の技術的な詳細や，CPU を初心者でも理解できるように工夫して説明がなされている記事が多数存在するため，多面的な説明が行われているといえる。本

[1]ChatGPT に指示を出すための命令文の名称。
[2]大規模言語モデルよりも規模が小さい言語モデルの名称。
[3]推論を行わせるプロンプティングの名称。

[4]全校（京都橘中学・高等学校）の生徒に対して行ったアンケートにおいて，590 人中 60 人が「CPU の説明がある程度できる」と回答した。

研究で使用した記事を表１にまとめた。

表１ 使用した記事

企業・サイト名	記事のタイトル
ロジテック INA ソリューションズ株式会社	CPU とは？概要や性能の見方を知ってパソコン選びに活かそう[6]
株式会社日経 BP	PC 自作の鉄則！2021[7]
にゃんさー	【中学生でもわかる】CPU とは？役割や性能の見方をわかりやすく解説[8]
PCBuildnet.com	【CPU とは？】自作 PC がグッと楽しくなる CPU の疑問集めました[9]

2.2 要約方法

本研究では，2.1.1 で紹介した記事から，「CPU とは」，「CPU の社名・ブランド」，「クロック周波数」の３つの題目に関連の近い文を抜き出し，複数文書をそれぞれ ChatGPT（GPT-3.5）に要約させる。３つの題目を抜き出した理由は，異なる題目での結果を検証するためである。

各題目の単語数について，**「CPU とは」は 408 単語，「CPU の社名・ブランド」は 602 単語，「クロック周波数」は 412 単語**であった。

次に紹介するプロンプトの始まりと終わりは「“　”」で挟む。また，プロンプトの改行をする部分には「¥n」と書く。３種の異なるプロンプトで要約・比較を行った。

I. **“(複数文書)”（プロンプトなし）**

生成に関する指示がない入力である。また，この入力においては各記事の区切りがない。

II. 〈段階１〉 **“##文書## ¥n （各文書に｛１｝～｛４｝まで番号を振った複数文書） ¥n ##指示## ¥n 文書から特に重要な単語１～５つを「,」で区切って生成してください。”**
→重要語句生成

〈段階２〉 **“##文書## ¥n （各文書に｛１｝～｛４｝まで番号を振った複数文書） ¥n ##重要語句## ¥n 【〈段階１〉で生成した重要語句】 ¥n ##指示## ¥n 重要語句を参考にして，文書をまとめてください。”**

〈段階１〉では，記事から重要語句を探すように指示する。このプロセスと近い研究が，教師あり Fine-tuning された LM を使用して要約のヒントとなる語句を生成している Li らの論文である。しかし，一般の人々がこのような LM を製作することは困難であるため，〈段階１〉では教師あり Fine-tuning された LM を使用せず，ChatGPT のみで語句を探し出すことを試みた。

〈段階２〉では，その重要語句を元に生成を行った。

また，「##文書##」と書くように，目的を強調させることで ChatGPT が理解しやすい形を取る事ができる[10]。

III. 〈段階１〉 **“##文書## ¥n （各文書に｛１｝～｛４｝まで番号を振った複数文書） ¥n ##指示## ¥n 文書から特に重要な単語１～５つを「,」で区切って生成してください。”**
→重要語句生成

〈段階２〉 **“##文書## ¥n （各文書に｛１｝～｛４｝まで番号を振った複数文書） ¥n ##重要語句## ¥n 【〈段階１〉で生成した重要語句】 ¥n ##指示## ¥n (1). 重要語句についての説明を行ってください。 ¥n (2). (1)をもとに文書をまとめてください。”**

〈段階１〉はⅡの〈段階１〉と同様である。

〈段階２〉では，重要語句についての理解を深めた後に，要約を指示している。このプロセスは，Wei らが考案した，連鎖的に思考を行う CoT プロンプティングを参考に行った。本研究では，重要語句を ChatGPT に思考させてから要約を行わせるという CoWT プロンプティングを新たに考えた。

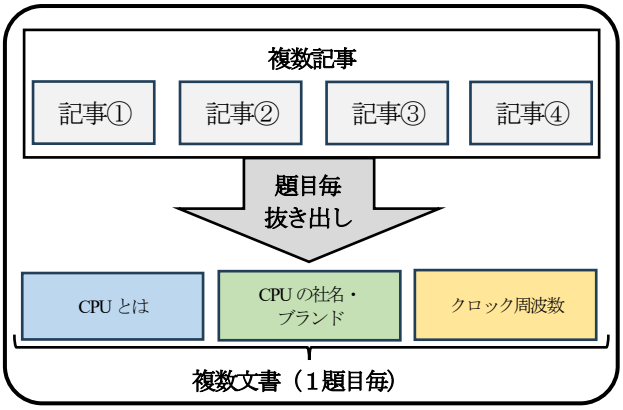

図２ 本研究の流れ（題目抜き出し）

2.3 解析方法

はじめに，生成された文書を，日本語の形態素解析器であり，日本語のテキストを形態素ごとに分割することができるツールである Janome[11]によって形態素解析（生成された文書を単語ごとに分割）を行った。

次に，形態素解析された文書について，どれほど簡潔かつ精確な要約が行われているのかについて評価した。生成された文書の圧縮率，ROUGE[12] を使用し，解析・評価を行うことで，信頼度の高い結果となる。

生成の安定性については，生成された単語数の分散を求めることで評価した。単語数の圧縮率や ROUGE，分散に関しては要約を１０回生成させることによって求めた。

圧縮率，ROUGE の統計解析には医薬学データ用統計解析プログラムの Steel-Dwass の多重検定を用いた（群数：3)。この検定を用いた理由は，生成に外れ値がある懸念や，生成が正規分布でない可能性を考慮したためである。

◆ ROUGE（ROUGE-N）について

ROUGE とは，Lin らによって開発された，要約システムの自動評価法として最も広く用いられており，要約の精度を測るものである。記事と要約した文書がどれだけ近似しているかを Precision（適合率）・Recall（再現率）で示すことができる。Lin らは，N（単語単位）を 1～4 まで変化させ，マニュアル評価

結果との相関を調べた結果，N=1,2 が最も高い相関であったと報告している。よって，本稿では 1 単語単位（ROUGE-1）で研究を行った。

★ **例として，記事が 20 単語，要約した文書が 10 単語，一致した単語が 5 単語であったとする。**

$\text{Precision} = \frac{\text{一致した単語数}}{\text{要約した文書の単語数}}$ よって $\frac{5}{10} = 0.50$ ．

$\text{Recall} = \frac{\text{一致した単語数}}{\text{記事の単語数}}$ よって $\frac{5}{20} = 0.25$ ．

一般的に，このスコアは 100 倍することから，Precision は 50，Recall は 25 となる。

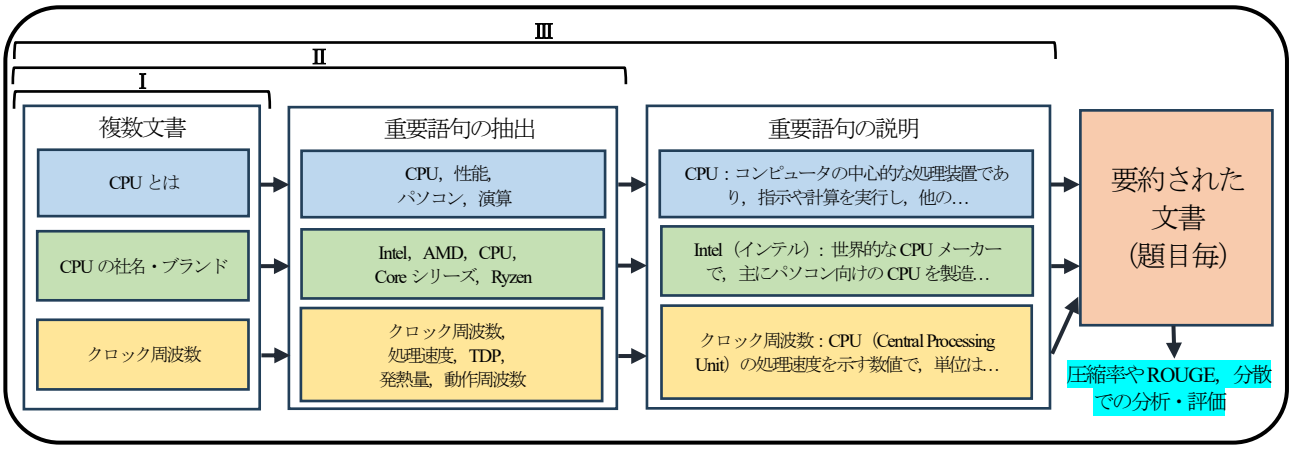

図３　本研究の流れ（プロンプト - 分析・評価）

3.　研究結果

3.1　簡潔性の評価

圧縮率が高いほど，簡潔に生成が行われているといえる。図４において，各題目のⅠ・ⅡとⅢの間に 5%水準で有意な差が認められた。一方で，各題目のⅠとⅡの間には有意な差が認められなかった。

したがって，Ⅲが最も圧縮率が高く，より簡潔な要約であるといえた。

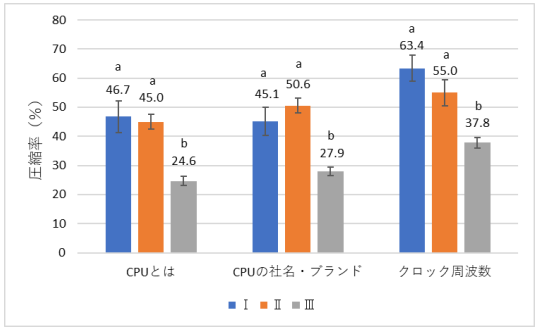

図４　文書の圧縮率（n＝10，エラーバー：標準誤差，$p<0.05$）

3.2　精確性の評価

ROUGE を利用して要約の精度を見ることができる。

Precision（図５）の「CPU とは」については，Ⅰ・ⅡとⅢに 5%水準で有意な差が認められた。また，「CPU の社名・ブランド」・「クロック周波数」についてはⅠとⅡ・Ⅲに有意な差が認められた。

Recall（図６）の「CPU とは」と「クロック周波数」について，Ⅰ・ⅡとⅢの間に有意な差が認められ，「CPU の社名・ブランド」についてはⅠとⅡ，ⅡとⅢの間に有意な差が認められた。

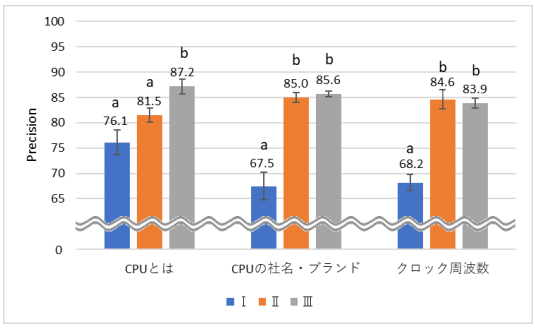

図５　文書の Precision（n＝10，エラーバー：標準誤差，$p<0.05$）

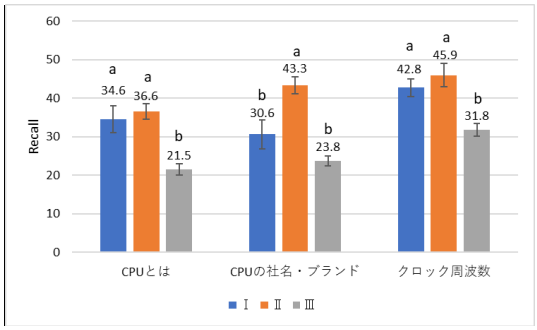

図６　文書の Recall（n＝10，エラーバー：標準誤差，$p<0.05$）

3.3　安定性の評価

単語数の分散が小さいほど，１生成ごとの単語数のばらつきが小さくなる。すなわち，図７では，ChatGPT による生成の安定性を確認することができる。Ⅰは生成に関する指示がないため，Ⅱよりもばらついた生成であった。Ⅱは，Ⅲよりも分散は少し大きいことがわかった。

よって，Ⅲは最も分散が少なく，安定した結果を出すことができているといえた。

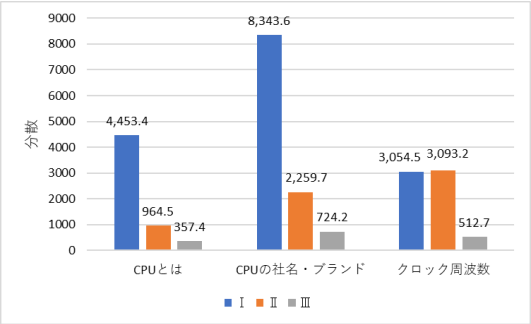

図 7 生成された文書の分散（n＝10）

4. 考察

3.1，3.3 から，Ⅲの CoWT プロンプティングが最も安定していて，文書を簡潔に要約することができていることがわかった。

Recall については，3.2 の分析結果から，ⅢがⅠやⅡに比べて低い結果にあった。Recall の計算式は $\frac{\text{一致した単語数}}{\text{記事の単語数}}$ であり，簡潔に要約するほど一致する単語数は減ることになり，Recall の値は小さくなる。一方，Precision の計算式は $\frac{\text{一致した単語数}}{\text{要約した文書の単語数}}$ であり，精確かつ簡潔に要約を行った際は一致した単語数が減るが，要約した文書の単語数も減るので Precision の値は高くなる。よって，Recall の値が低いかつ Precision の値が高い場合は精確に要約ができていると判断できる。

5. 結論

インターネット上の情報は膨大であり，同一語句について説明する記事は数多く存在する。これらの記事について ChatGPT を使って合成し，多面的な情報に触れることができる要約を行いたいと考えた。よって本稿では，複数の CPU についての記事を元に，ChatGPT（GPT-3.5）を利用し，簡潔かつ精確で安定性のある複数文書要約を行うためのプロンプトを研究した。

大規模言語モデルによる文書生成では，プロンプト（指示）が文書生成の内容に大きく影響するため，さまざまなプロンプトが提案されている。プロンプトにおいて思考連鎖を行わせる手法や，特に文書要約において要約のヒントとなる語句を元に要約を行わせる手法が提案されている。それゆえ本稿では，明確な指示とともに重要語句の思考連鎖によって要約を行わせる“Chain-of-Words-Thought prompting”（CoWT プロンプティング）を考案した。これによって，より簡潔で精確かつ安定的に複数文書の合成が可能になると考えた。

研究にあたって，生成した単語数の分散，生成された文書の圧縮率，ROUGE（要約システムの自動評価法）である Precision（適合率）や Recall（再現率）から生成結果を判断することを試みた。

CPU という題材においては，CoWT プロンプティングによって出力された要約文の簡潔性，精確性，安定性が最も高かった。複数文書要約における CoWT プロンプティングの有用性が示された。

6. 今後の展望・課題

今後も AI の技術を利用した文書の要約は盛んになっていくだろう。本研究において CPU という題材においては CoWT プロンプティングの有用性を確かめることができた。しかし，題材の数が少ないことは明らかであり，CPU 以外の複数の題材においての検証は必ず行う必要がある。

本研究は，複数文書中に不精確な情報は含まれないことを前提としていた。しかし，複数文書中には，反対の主張が含まれている場合や不精確な情報が含まれている可能性は十分にある。こうしたイレギュラーな事態にも対応可能な思考を行うプロンプトの研究も今後の大規模言語モデルを用いた複数文書要約の研究に求められることだと考える。

謝辞

本研究にあたって，記事の使用を快く承諾してくださいましたロジテック INA ソリューションズ株式会社様，株式会社日経 BP 様，にゃんさー運営のがーと様，PCBuildnet.com 運営の陣内聡様には深く感謝の意を表します。

参考文献

(1) 西川 仁：“深層学習による自動要約”，人工知能，34 巻 4 号（2019）.
(2) OpenAI：“ChatGPT”（GPT-3.5），https://chat.openai.com/（2023/12/29 閲覧）.
(3) Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao，Xifeng Yan：“Guiding Large Language Models via Directional Stimulus Prompting.”，arXiv:2302.11520（2023/10/9）.
(4) Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, Denny Zhou：“Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”，arXiv:2201.11903（2023）.
(5) Songlin Bao，Tiantian Li，Bin Cao：“Chain-of-event prompting for multi-document summarization by large language models”，International journal of Web information systems（2024）.
(6) ロジテック INA ソリューションズ株式会社：“CPU とは？概要や性能の見方を知ってパソコン選びに活かそう”，https://www.pro.logitec.co.jp/houjin/usernavigation/hddssd/20210507/（2023/12/31 閲覧）.
(7) 滝 伸次，影山 巧：“PC 自作の鉄則！2021”，pp.18-31，日経 PC21 編・日経 BP（2021）.
(8) がーと：“【中学生でもわかる】CPU とは？役割や性能の見方をわかりやすく解説”，https://nyanswer.com/whats-cpu/（2023/10/1 閲覧）.
(9) 陣内聡：“【CPU とは？】自作 PC がグッと楽しくなる CPU の疑問集めました”，https://pcbuildnet.com/about-cpu/（2023/10/1 閲覧）.
(10) DAIR.AI：“General Tips for Designing Prompts”，https://www.promptingguide.ai/introduction/tips（2023/12/30 閲覧）.
(11) 打田智子：“Janome”，https://mocobeta.github.io/janome/（2023/12/30 閲覧）.
(12) Lin, C.-Y：“ROUGE：A package for automatic evaluation of summaries”，Proc. Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL 2004（2004）.