

生成 AI を用いた単語間における類推

田代寛治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

m23j4028u@edu.tottori-u.ac.jp

murata@tottori-u.ac.jp

概要

「発想や類推を機械によって行う、あるいは人間の類推の補助が出来ないか」というテーマで研究を行う。例として (爆発物) - (ガソリン) → (水素) からエンジンの代替エネルギーの発想を得るなどである。

以前の村田の研究 [1] では、「ネコ」という単語の属性を考え、その属性に「大きい」という属性を追加することで「トラ」という単語を類推して、「ネコを大きくするとトラが導き出される」拡大や縮小の類推を行った。田代の研究 [2] では、類似する 2 単語の引き算により得られる属性を類推する研究を行った。これらの研究では McRae のデータセット [3] を基に研究を進めた。

本研究は、入力次第で様々な類推が可能な ChatGPT を使用し、「田代の研究で行った類推との比較」、「McRae と同様のデータセットの作成」、「ChatGPT による様々な類推」を行った。田代の研究の引き算の類推では類推性能 0.46 であったが、本研究の引き算の類推では類推性能 0.48 に向上した。ChatGPT で引き算の類推を行う時も集合で引き算をした場合の方が向上することが分かった。また、McRae のようなデータセット生成で意味の一致を正解とした場合、性能 F 値 0.61 と高性能なデータセット作成が可能であることが分かった。9 種類のオズボーンのチェックリスト [4] での ChatGPT を用いた類推において、類推性能 0.88 で類推できることを確認できた。

1 はじめに

類推や発想の先行研究として村田の研究 [1]、須ヶ牟田の研究 [5]、田代の研究 [2]、梅村の研究 [6]、國藤の研究 [7]、Kenneth の研究 [8] などがある。これらの研究が行われた時は生成 AI が普及しておらず、使用されていなかった。そこで生成 AI の使用を検討し

た。近年普及している生成 AI の中でも、ChatGPT は膨大なテキストデータを用いた事前学習を経てさまざまな言語タスクに対する適応的な応答生成を可能にしている。特にユーザからの入力に対し、文脈を考慮した一貫性のある文章を生成する能力を持つ。この特徴を活かせると考え、本研究の類推には ChatGPT を使用した。

田代の研究 [2] では、類似した単語同士における類推を行い、類推を引き算式と見立てて式変形による類推方法の増加を目指した。この研究も McRae のデータセット [3] を基に研究を進めた。しかし、このデータセットだけではオズボーンのチェックリスト [4] にある類推手法を全て行うには不向きであった。そこで、本研究では ChatGPT を用いてデータセットの作成、ChatGPT による類推の実行を試みた。

2 先行研究

2.1 先行研究で用いたデータセット

McRae のデータセット [3] は 541 種類の単語に対して 725 名の評定者が与えた総計 7,259 の属性を収集したものである。この属性と単語のデータセットを用いて研究を行う。度数分布 (票数) とは 30 人に対して与えられた単語から何人が属性を発想できたかを示す値である。データセットの例を表 1 に示す。

2.2 先行研究の説明

田代の研究 [2] では、票数を割合に換算した票数比率と、内積を利用して類似しているかを判断する cos 類似度を用いて、4 通りの手法で引き算の類推をした。4 手法の概要として、表 1 のデータを用いて例示を行う。

例えば単語 A 「cheetah」から単語 B 「tiger」を引いて、cheetah のみが持つ属性 C 「is_fast」を類推、つまり A-B → C の類推をする。この時、2 単語は共通

表 1 McRae のデータセットの例

単語	属性	度数分布 (票数)
cheetah	is_fast	22
cheetah	has_black_spots	21
cheetah	an_animal	19
...
tiger	has_stripes	22
tiger	a_carnivore	21
tiger	an_animal	19
...

の属性 `an_animal` の票数 19 などの \cos 類似度で類似した単語組として選出される。類似した単語同士における引き算の方が、全く異なる単語同士における引き算よりも有意義な類推と考え、この選出を実行した。次に、選出された 2 単語のうち引かれる単語 `cheetah` のみが持つ属性を類推する際、票数が大きい属性 `is_fast` が類推される。属性 `is_fast` が、`tiger` には含まれない `cheetah` 固有の属性として求められる。票数が大きい場合は、固有な属性が類推され、票数が小さい場合は、単語同士において新しい発見に繋がる可能性がある属性が類推される。このような単語組と属性の選出における工夫を 4 種類実行し、4 手法とした。A 節に詳細を述べる。

3 先行研究の問題点

- オズボーンのチェックリスト [4] の全てに対応していない。
- McRae のデータセット [3] に依存し過ぎている。

4 タスクと提案手法

4.1 タスク 1

タスク 1 は、単語 A から単語 B を引いて思いつく属性を求めることである。例えば、`cheetah` から `tiger` を引いて `is_fast` を思いつくというものである。田代の研究 [2] と ChatGPT の 2 手法の合計 3 手法と比較する。ChatGPT の 2 手法は ChatGPT-a, ChatGPT-b とする。

ChatGPT-a は、ChatGPT により、類推で使う単語の McRae のデータセット [3] と同様の単語と属性からなる構造のデータを抽出してから単語に属性 (要素) が含まれるか含まれないかという集合の考え方で引き算を実行して類推を行う。

- 入力：壺に当てはまる要素とバケツに当てはま

る要素をそれぞれ取得して、壺のみが持つ要素を単語で教えて

- 出力：壺に当てはまる要素: 蓋がある, 液体を保持できる, 装飾的なデザインがある. バケツに当てはまる要素: 取っ手がついている, 比較的大容量である, 通気孔があることがある. これらの中で、壺のみが持つ要素は、「装飾的なデザインがある」

ChatGPT-b は、ChatGPT により、引き算の類推を直接行う。

- 入力：壺には当てはまっていてバケツには当てはまらない壺固有のものを類推して単語で出力
- 出力：「蓋 (ふた)」

4.2 タスク 2

タスク 2 は、McRae のデータセット [3] のような単語と属性を自動生成することである。タスク 2 の手法では、McRae のデータセットにあるような単語と属性のデータセットを 5 データセット `few shot` として ChatGPT に例示して、1 データセット出力させる。

- 入力：マクレーのデータベースでは、英単語に対して、英語で属性と属性の頻度 (出現しやすさ) を与えている。以下に、マクレーのデータベースの例を示している。これと同じような形式で、`hose` のマクレーのデータベースを作成せよ。
`hamster is_small 24 hamster has_fur 23 hamster a_pet 21...`
- 出力: `hose is_used_for_watering 20 hose is_long 18 hose is_made_of_rubber 17...`

4.3 タスク 3

先行研究ではオズボーンのチェックリストの一部しか扱えなかった。タスク 3 は、オズボーンのチェックリストの 9 項目すべての類推を ChatGPT で行うことである。以下に例を示す。

- 入力：オズボーンのチェックリストの「拡大・縮小する」の例として「拡大・縮小するもと」、「拡大・縮小した結果」、「拡大と縮小のどちらをしたか」を書いてください。
- 出力：もとのもの: 普通のスマートフォン
拡大: スクリーンサイズを大きくする結果: 動画やゲームに特化した「タブレット型スマホ」拡大を適用

縮小: スクリーンサイズを極小化し、腕時計サイズにする結果: スマートウォッチのような「ミニマムスマホ」縮小を適用

5 実験

5.1 実験に用いた ChatGPT のバージョン

タスク 1 で使用する ChatGPT の種類は GPT-3.5, タスク 2 で使用する ChatGPT の種類は GPT-4o-mini, タスク 3 で使用する ChatGPT の種類は GPT-4o である.

5.2 実験の評価方法

タスク 1 とタスク 3 は先行研究 [2] と同様に評価は人手で行う. それぞれ類推結果を 4 段階で評価して正解率を求める. 表で結果を示す時の「◎, ○」と「◎, ○, △」は複数の評価を 1 つの基準でまとめて正解率を求めた. 例えば「◎, ○」は「◎だけでなく, ◎と評価した人と○と評価した人の合計人数を分子, 評価人数の全体を分母」として正解率を求めた. 評価基準の 4 段階の内訳は以下のようにする.

- ◎: 類推できる
- ○: 大方類推できる, 少し変だと思う
- △: 見方によっては類推できる
- ×: 類推できない

タスク 2 の評価は McRae のデータセット [3] を正解とした時の ChatGPT の出力結果の評価として, 再現率 (Recall), 適合率 (Precision), F 値をデータセットごとに計算する. 意味の一致や適切な属性に関しては, 人手で判断した. それぞれの数値を以下に示す.

- 完全一致した属性の種類の数と重みとして求めた再現率 1, 適合率 1, F 値 1
- 完全一致した属性の票数を重みとして求めた再現率 2, 適合率 2, F 値 2
- 意味は一致した属性の種類の数と重みとして求めた再現率 3, 適合率 3, F 値 3
- 意味は一致した属性の票数を重みとして求めた再現率 4, 適合率 4, F 値 4
- 単語に対して適切な属性の種類の数と重みとして求めた適合率 5
- 単語に対して適切な属性の票数を重みとして求めた適合率 6

重みに票数も考慮したのは, 票数が多い属性で正解している方が望ましいと考えたためである. 適合率 5 と適合率 6 は ChatGPT の出力結果がそもそも McRae のデータセットに関係なく, 正解であるかを人手評価した数値であるため, 適合率のみを求めた. A を正解とする McRae のデータセットの属性の集合 (種類の数と票数), R を ChatGPT の出力結果に

含まれる属性の集合 (種類の数と票数) とすると, これら 3 つの値は次のようになる.

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

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

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

5.3 タスク 1 の結果

先行研究の 4 手法の類似度合などの指標値の上位 5 組, つまり合計 20 組で行った. 被験者は著者を除く 5 名である. 手法 X により得られるデータをデータ X と呼ぶ. 先行研究の手法では, データ X には手法 X を用いた.

タスク 1 の被験者実験の結果を表 2, 表 3, 表 4 に示す.

表 2 タスク 1 の先行研究の類推評価

	◎	◎, ○	◎, ○, △	平均
データ 1	0.16	0.36	0.56	0.36
データ 2	0.28	0.52	0.80	0.53
データ 3	0.44	0.68	0.76	0.63
データ 4	0.12	0.28	0.48	0.29
データ平均	0.26	0.47	0.65	0.46

表 3 タスク 1 の ChatGPT-a の類推評価

	◎	◎, ○	◎, ○, △	平均
データ 1	0.24	0.36	0.60	0.40
データ 2	0.44	0.52	0.52	0.49
データ 3	0.60	0.84	0.88	0.77
データ 4	0.16	0.32	0.60	0.36
データ平均	0.34	0.48	0.63	0.48

表 4 タスク 1 の ChatGPT-b の類推評価

	◎	◎, ○	◎, ○, △	平均
データ 1	0.08	0.20	0.44	0.24
データ 2	0.12	0.36	0.56	0.35
データ 3	0.24	0.44	0.72	0.47
データ 4	0.08	0.36	0.44	0.29
データ平均	0.13	0.34	0.53	0.33

5.4 タスク 1 の考察

先行研究の手法での指標値の上位データを利用して実験を行ったもので, 先行研究に有利な評価実験となっている. 先行研究の方が有利にも関わらず, ChatGPT-a の方が類推性能が 0.48 で, 先行研究の類推性能 0.46 を上回った. ChatGPT-b の性能が ChatGPT-a より下がったことから, ChatGPT-a の方法のように集合と要素を用いた引き算を入力で指定する方が高性能になることが判明した.

5.5 タスク 2 の結果

30 セット出力し評価した。人手評価はそのうちの 5 セットで行った。タスク 2 の実験結果を表 5, 表 6 に示す。タスク 2 の出力の一部を表 7, 表 8 に示す。表 8 の不一致は, 表 7 に is_heavy と類似した属性が無いという意味である。

表 5 タスク 2 の F 値と再現率と適合率の平均値

平均 F 値 1	平均再現率 1	平均適合率 1
0.22	0.27	0.18
平均 F 値 2	平均再現率 2	平均適合率 2
0.25	0.30	0.22
平均 F 値 3	平均再現率 3	平均適合率 3
0.51	0.60	0.46
平均 F 値 4	平均再現率 4	平均適合率 4
0.61	0.69	0.55

表 6 タスク 2 の適合率 5 と適合率 6 の平均値

平均適合率 5	平均適合率 6
0.99	0.99

表 7 タスク 2 の McRae のデータセットの例

単語	McRae 概念属性	McRae 度数分布 (票数)
hose	is_long(完全一致)	20
hose	is_round(意味は一致)	8
hose	is_green	15
...

表 8 タスク 2 の ChatGPT の出力の例

単語	ChatGPT 概念属性	ChatGPT 度数分布 (票数)
hose	is_long(完全一致)	18
hose	can_be_coiled(意味は一致)	12
hose	is_heavy(不一致)	5
hose	is_flexible(適切な属性)	16
...

5.6 タスク 2 の考察

タスク 2 は最初にデータセットを例示したとはいえ, F 値 4 が 0.61 という比較的高い性能でデータセットの作成が行えた。また, 適合率 5 と適合率 6 も高性能であることから, McRae のデータセット [3] と同様の正しいデータが格納されたデータセットの作成が可能であることを確認できた。これにより, ChatGPT が間違っただけのデータセットを作成する可能性は低く, 人手でなく生成 AI によって正確なデータセットの作成が可能であると思われる。将来的には, この種のデータを自動構築し, 様々な研究に役立てることも考えられる。

5.7 タスク 3 の結果

9 項目全てにおいて類推 5 個ずつ 45 組実行し, 4 段階で評価した。オズボーンのチェックリスト [4] の 9 項目のそれぞれのタスク 3 の被験者実験の結果を表 9 に示す。

表 9 タスク 3 の 9 項目の類推性能

類推の種類 (項目)	◎	◎, ○	◎, ○, △	平均
1.Substitute	0.80	1.00	1.00	0.93
2.Combine	0.76	0.96	1.00	0.91
3.Adapt	0.84	1.00	1.00	0.95
4.Modify	0.76	0.88	1.00	0.88
5.Put to Another	0.64	0.84	0.96	0.81
6.Eliminate	0.84	0.96	1.00	0.93
7.Reverse	0.48	0.76	0.92	0.72
8.Rearrange	0.64	0.84	1.00	0.83
9.Magnify/Minify	0.88	0.92	0.96	0.92
9 項目の平均	0.74	0.91	0.98	0.88

5.8 タスク 3 の考察

1 項目あたりの個数が少ないせいで上振れた可能性もあるが, 高性能であった。特に平均性能が低かった 7.Reverse に関しても 0.72 とタスク 1 の ChatGPT-a よりも高性能であった。これにより, McRae のデータセット [3] に依存しない, ChatGPT を用いた類推によって, 間違っただけの類推や見当違いな類推を行う可能性は低いと考える。

6 おわりに

McRae に有利な条件で行った引き算の類推で, 類推性能 0.46 に対して ChatGPT の類推性能 0.48 と高い性能での類推が可能であることが確認できた。また, ChatGPT で引き算の類推を行う時も集合で引き算をした場合の方が向上することが分かった。McRae のデータセット [3] の作成も意味の一致を正解にするなら, F 値 0.61 と高性能で行えることが分かった。また, 9 種類のオズボーンのチェックリスト [4] での ChatGPT の類推では 0.88 と高性能な類推を行わせることが可能であることを確認できた。今後は, 本稿のタスク 3 の実験では, 単語を入力としてその単語に関わる類推を出力する実験を行いたい。また, オズボーンのチェックリスト以外にも様々な類推法がある。さらに多くの種類の類推法を ChatGPT で試したい。

参考文献

- [1] Masaki Murata, Idea Generation Support Using Conceptual Attributes, Information Engineering Express, Vol.9, No.2, pp1-10, 2023.
- [2] 田代寛治: 属性の頻度で選出した類似単語において引き算を用いた類推, 鳥取大学卒業論文, 2023.
- [3] Ken McRae, George S Cree, Mark S Seidenberg, and Chris McNorgan, Semantic feature production norms for a large set of living and nonliving things. Behavior Research Methods, 37, pp.547-559, 2005.
- [4] Osborn, Alex F, Applied imagination : principles and procedures of creative problem-solving. New York : Scribner, 1957.
- [5] 須ヶ牟田稔: 概念の結合に基づく類推を利用した発想支援, 鳥取大学卒業論文, 2021.
- [6] 梅村雄貴, アイデア連想のつながりを明示化するブレインライティング支援システム, 北陸先端科学技術大学院大学修士論文, 2019.
- [7] 國藤進, 発想支援システムの研究開発動向とその課題, 人工知能学会誌, 8 巻, pp.552-559, 1993.
- [8] Kenneth Ward Church. Word2vec. Natural Language Engineering, 23(1):155–162, 2017.

A 先行研究の詳細

田代の研究 [2] では、票数を割合に換算した票数比率と内積を利用して類似しているかを判断する cos 類似度を用いて、4 通りの手法で類推した。

- 手法 1: 票数ありの属性の cos 類似度による単語組選出と票数比率による属性選出

→単語 A と単語 B の cos 類似度と単語 A での属性 C の票数比率の積が大きい 3 つ組を抽出する。単語 A と単語 B のベクトルは属性の票数の値をベクトルの要素の値として表現し、cos 類似度を算出する。属性 C は単語 A の属性のうち単語 B の属性でないものとする。属性 C の票数比率は、単語 A での属性 C の票数を属性の単語 A の属性の票数の和で割ったものである。

- 手法 2: 票数ありの属性の cos 類似度による単語組選出と乱数による属性選出

→単語 A と単語 B の cos 類似度が大きい 3 つ組を抽出する。単語 A と単語 B のベクトルは属性の票数の値をベクトルの要素の値として表現し、cos 類似度を算出する。属性 C は単語 A の属性のうち単語 B の属性でないものとする。

- 手法 3: 票数なしの属性の cos 類似度による単語組選出と票数比率による属性選出

→単語 A と単語 B の cos 類似度と単語 A での属性 C の票数比率の積が大きい 3 つ組を抽出する。単語 A と単語 B のベクトルは属性をベクトルの要素とし、要素の値は全て 1 に固定して表現して cos 類似度を算出する。属性 C は単語 A の属性のうち単語 B の属性でないものとする。属性 C の票数比率は、単語 A での属性 C の票数を属性の単語 A の属性の票数の和で割ったものである。

- 手法 4: 票数なしの属性の cos 類似度による単語組選出と乱数による属性選出

→単語 A と単語 B の cos 類似度が大きい 3 つ組を抽出する。単語 A と単語 B のベクトルは属性をベクトルの要素とし、要素の値は全て 1 に固定して表現し、cos 類似度を算出する。属性 C は単語 A の属性のうち単語 B の属性でないものとする。

B タスク 2 の入出力の例

- 入力：マクレーのデータベースでは、英単語に対して、英語で属性と属性の頻度（出現し

やすさ）を与えている。以下に、マクレーのデータベースの例を示している。これと同じような形式で、hose のマクレーのデータベースを作成せよ。hamster is_small 24 hamster has_fur 23 hamster a_pet 21 hamster lives_in_cages 21 hamster an_animal 19 hare a_rabbit 20 hare has_fur 19 hare has_ears 17 hare is_fast 16 hare beh_hops 14 harpoon is_sharp 12 harpoon made_of_metal 11 harpoon used_for_hunting 10 harpoon used_for_killing_whales 10 harpoon a_weapon 8 hatchet made_of_metal 15 hatchet has_a_handle 11 hatchet used_for_cutting 11 hatchet a_tool 9 hatchet used_for_chopping 8 hornet beh_stings 24 hornet beh_flies 19 hornet has_wings 17 hornet an_insect 10 hornet lives_in_a_nest 10

- 出力：こちらが「hose」のマクレーのデータベース形式の例です：hose is_used_for_watering 20 hose is_long 18 hose is_made_of_rubber 17 hose is_flexible 16 hose has_a_nozzle 15

C タスク 3 の入出力の例

- 入力：オズボーンのチェックリストの Modify（修正する）の例として、修正前のもの、そのうちの修正されるもの、どのように修正されるか、修正された結果できるもののセットを 5 個書いてください。
- 出力：製品の形状を修正修正前: 四角形の弁当箱修正されるもの: 弁当箱の形状修正方法: 角を丸くして楕円形にする修正後: 持ちやすくなり、子どもでも扱いやすい丸い弁当箱