

代替用途テストの柔軟性自動評価手法に関する検討

深田 龍之介[†] 上田 樹美[†] 石井 裕剛[†] 下田 宏[†] 大林 史明[†]
京都大学[†]

1 はじめに

近年、創造的思考が重要視されている。創造的思考は、一つ概念や考えから多種多様なアイデアを生み出す発散的思考と、多種多様なアイデアから一つの結論となるアイデアに収束させていく収束的思考の二つの思考に大別される [1]。発散的思考のパフォーマンスを向上させることは、創造的思考のパフォーマンスを向上させることに繋がるため、発散的思考を支援する手法が多く提案されてきた。

発散的思考を支援する手法を評価する代表的な手法の一つに代替用途テスト (Alternative Uses Test、以降 AUT) [2] がある。AUT は提示されたモノの本来とは異なる使い方のアイデアをできるだけ多く回答するタスクである。以降提示されるモノを「題目」と表記する。発散的思考のパフォーマンスの評価指標には流暢性、柔軟性、独創性などがあり [3]、AUT を評価する際は各指標のスコアとしてアイデアの個数や種類数、希少度合いを評価する。

従来、AUT での各指標のスコアの算出に際して実験結果を人手で評価する必要があるが、評価者の経験や知識に基づく主観が結果に大きな影響を与えてしまう可能性があること、また膨大な労力や時間がかかってしまうことが問題点として挙げられる。そこで本研究では、特に発散的思考の柔軟性に注目し、AUT の回答から自動で算出できる、発散的思考の柔軟性評価のための指標について検討した。人を対象とした実験で AUT の回答を収集し、従来の人手による評価を実施したうえで、得られた回答に対して自然言語処理の手法を適用して得た各回答の分散表現に基づき柔軟性の新たな指標を算出し、従来の人手により評価された柔軟性スコアとの比較を行う。

2 実験

本章では AUT の回答を収集するために実施した実験、および従来手法に基づき実施した評価について述べる。

2.1 実験手順

実験は 2021 年 12 月 24 日から 2022 年 1 月 12 日の期間に実施した。実験は 1 人ずつ、合計で 32 人が行った。参加者は実験に関する説明を受けたのち、同意の上で実験に参加した。例題として 2 つの題目で AUT を行い、さらに本題として 4 つの題目で AUT を行った。回答の制限時間は例題、本題どちらも 1 つの題目あたり 4 分間とし、各題目間に 45 秒の休憩時間を設けた。提示する題目には例題で「段ボール箱、スポンジ」、本題では「割り箸、ペットボトル、縄跳び、下敷き」を用いた。また本題では題目の提示順による影響を排除するため、題目の提示順に関してカウンターバランスをとった。AUT 終了後、提示した題目の主観的な難易度に関するアンケートを実施した。なお、本実験は京都大学大学院エネルギー科学研究科の「人を対象とする研究倫理委員会」の承認を得て実施した。

2.2 評価手法

本実験では発散的思考の柔軟性の指標として、AUT の柔軟性スコアを評価した。以下では AUT の柔軟性スコア評価手法について述べる。まず実験にて回答された各アイデアが「本来とは異なる使い方」になっているか等、回答として適切かどうかを評価した。次に適切であると評価されたアイデアに関して、表 1 に示す 3 つの観点から分類し、この分類が全て同じであるアイデアを同じ種類のアイデアと判断した。このアイデアの種類数を AUT の柔軟性スコアとした。以上の評価は第一著者と、著者と同じ研究室に属する学生の合計 3 名で実施し、全員が独立に評価を行ったのちに、異なる評価をしたアイデアに関して議論を行った上で最終的な評価を決定した。これらの評価の結果、32 名の参加者から 754 の有効な回答を得た。

A Study on a Method for Automated Evaluation of Flexibility of Alternative Uses Test

[†] Ryunosuke Fukada, Kimi Ueda, Hirotake Ishii, Hiroshi Shimoda, Fumiaki Obayashi, Kyoto University

表1 アイデアを分類する観点

- | | |
|----|--|
| 1: | そのアイデアは題目を「そのまま」使っているか、または「材料」として使っているか。 |
| 2: | そのアイデアはどのようなモノの代替、または模倣として使っているか。 |
| 3: | そのアイデアは提示された題目をどのように使っているか。 |

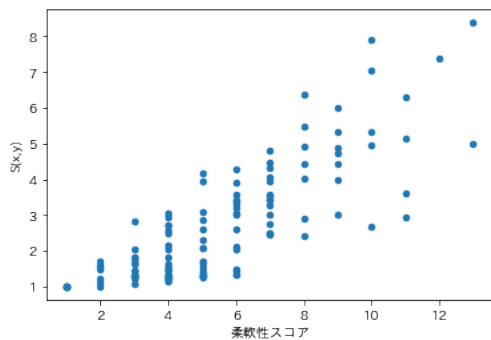


図1 AUT 柔軟性スコアと $S(x, y)$ の相関

3 検討した自動評価手法

本章では自然言語処理を用いた柔軟性の新たな指標の算出手法について述べる。上記で述べたように AUT の柔軟性スコアは回答の種類数を表している。そこで各参加者が新しくアイデアを回答した際、すでに回答したアイデアとの非類似度を数値化することで種類数を表現することができるのではないかと考えた。

まず、入力する回答を各参加者が回答した順に並べ替えた。次に各回答を形態素解析した。形態素解析器には Sudachi[4]、辞書には SudachiDict を用いた。次に各回答から分散表現を得た。分散表現の獲得には日本語 wiki コーパスをもとに学習した Doc2Vec[5] を使用した。その後コサイン類似度をもとにして式 (1) に従い $S(x, y)$ を算出した。ただし参加者 x が題目 y に関して i 番目に回答したアイデアと $1 \sim i - 1$ 番目までに回答していたアイデアとの最大のコサイン類似度を $s_{max}(x, y, i)$ とおく。また $s_{max}(x, y, 1) = 0$ とした。

$$S(x, y) = \sum_{i=1}^n (1 - s_{max}(x, y, i)) \quad (1)$$

4 結果と考察

人により評価した AUT の柔軟性スコアを横軸に、3章の手法により算出した $S(x, y)$ を縦軸にプロットしたものを図1に示す。これら2つの指標に関してピアソンの相関係数は0.816、 p 値は 8.77×10^{-32} であり、有意な相関が見られた。この結果から、従来の柔軟性スコアと相関のある指標を人手を使わずに算出することができると可能性が示唆された。

評価精度を高めるための改善点として、さらにデータ数を増やし学習を行うこと、分散表現を作成する際に BERT[6] のような別の手法を適用すること等を検討している。またコサイン類似度は文章が長いほど類似度が高くなる傾向を考慮し、AUT 出題時に回答の長さや記載方法についての教示を加えることも検討している。

参考文献

- [1] Guilford, J. P.: The nature of human intelligence. (1967).
- [2] Wilson, R., Christensen, P., Merrifield, P. and Guilford, J.: Alternate Uses-Form A, *Manual of administration, scoring and interpretation*. Beverly Hills, Calif.: Sheridan Supply Corp (1960).
- [3] Torrance, E. P.: The nature of creativity as manifest in its testing, *The nature of creativity*, pp. 43-75 (1988).
- [4] Takaoka, K., Hisamoto, S., Kawahara, N., Sakamoto, M., Uchida, Y. and Matsumoto, Y.: Sudachi: a Japanese Tokenizer for Business, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France (2018).
- [5] Le, Q. and Mikolov, T.: Distributed representations of sentences and documents, *International conference on machine learning*, PMLR, pp. 1188-1196 (2014).
- [6] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).