

類推能力の評価方法としての二項が未知の四項類推課題の提案

Two-terms-missing-in-four-terms Analogy as a “Right” Task to Evaluate Analogy Competence

加藤 龍彦[†], 日高 昇平[†]
Kato, Tatsuhiko, Hidaka, Shohei

[†]北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology
skylark@jaist.ac.jp

概要

対象間の関係の関係性を用いる推論を類推という。近年提案された単語埋め込みモデルは類推課題に高精度で回答可能であり、類推をモデル化する上で新たな可能性を示している。しかしこうしたモデルを評価するのに用いられてきた課題は、関係の関係性を用いなくても正答できる可能性がある。本稿では関係の関係性を用いなければ正答できない課題を提案し、単語埋め込みモデルがこうした課題をほとんど解けないことを示すことで、モデルが類推を行っていない可能性を示した。

キーワード：類推, 四項類推, 単語埋め込みモデル

1. 類推～関係の関係性についての推論

「男」と同じ関係を持つ「女」という語に対して、「王」と同じ関係を持つ語は何か」というように、「関係の間の関係性」の理解を必要とする課題を解く能力を類推という(Gentner, 1983; Gentner, 2010)。この例では「男女」の関係性と、王と何らかの単語の関係性が「同じ」であるというところから、その語が「女王」である、という推論が可能になる。類推能力は未知の対象や、まだ完全にはわかっていない事柄について推測することを可能にすることで、科学的発見、教育や学習などの場面で欠かせない能力だと考えられてきた(Holyoak and Thagard, 1989)。

こうした類推能力をモデル化する試みは認知科学において早くから行われてきたが、ほとんどのモデルはモデラーが対象間の関係などの意味の表現を明示的に与えることで作成されていた(Falkenhainer et al., 1989; Hummel and Holyoak, 2003)。これに対して、近年自然言語処理分野で提案された単語埋め込みモデルとして知られる word2vec (Mikolov, 2013a,b) や GloVe (Pennington et al., 2014)といったモデルは、意味の表現

自体を言語データ中の共起確率を近似することで学習し、なおかつ類推課題に高い精度で回答できることが示されている。また単語埋め込みモデルは意味判断においても人と類似した判断を行うことが知られており(Baroni et al., 2014)、モデルを通じて人の意味判断や類推の特性を明らかにできる可能性を示している。

単語埋め込みモデルの類推課題としては、最初にあげたような3つの単語からなる問いに答える四項類推課題が典型的に用いられる。こうした課題はモデルが単語間の意味関係をどの程度良く表現できているかを検証するために使われる。この四項類推課題に対し、単語埋め込みモデルでは、課題の3単語のベクトルの和・差 $v_c - v_b + v_a$ と候補単語 v_d の余弦類似度を最大にする単語 $v = \operatorname{argmax}_{v_d} \cos_{\{v_a, v_b, v_c\} \in V}(v_c - v_b + v_a, v_d)$ を類推結果として回答する。この式は、類推課題中の単語 a,b,c とベクトル空間上で「平行四辺形」

(図1右)の配置にある単語と最も類似した単語を回答すると想定されている。もし word2vec モデルが本当にこうした平行四辺形関係を捉えられているのであれば、単語ベクトル v_a と v_b, v_c と v_d の間の差分としての関係が「同じ」である、ということ捉えていることになる。これはある意味での関係の関係性を用いており、その意味で類推を行っていると言ってよい。

本稿では、このような四項類推課題を用いてモデルの類推能力を測るには本質的な限界があることを指摘し、類推能力を評価する上でより適切な課題として「二項が未知の四項類推」(以下、二項未知類推と略記)を提案する。さらに、代表的な単語埋め込みモデルである word2vec では二項未知類推課題に十分な精度で回答できないことを示し、この課題に回答できるモデルを構築するための方策を提案する。

2. 既存の四項類推の問題点と二項が未知の四項類推

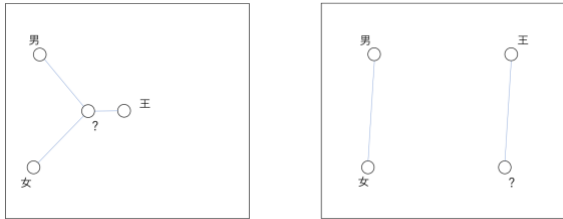


図 1: 通常の四項類推課題では、左のような関係の同一性に基づかないような単語対でも正答としてしまう

従来の(一項が未知の)四項類推課題の問題点は、問いにある 3 単語のそれぞれと候補単語の関係を利用することで、「関係性の関係」に関する推論を行わずに課題に回答可能な点である。例として、最初にあげた男:女::王:?の問いに対しては、「男」、「女」、「王」それぞれの単語と類似あるいは非類似であるという情報の利用だけでも正解と想定される「女王」を回答可能となる可能性がある(図 1 左)。実際 Rogers ら(2017)は、類推課題における word2vec の正答率が、問題文にある単語(e.g., 前出の例では「王」と正解単語の類似度に依存することを報告している。このことは、word2vec が単語ベクトル間の関係の関係性ではなく、それらの関係だけで四項類推課題に回答していることを示唆する。このように通常の四項類推課題では関係の関係性以外の情報を用いて正答を得ることが可能であるため、必ずしも類推能力を評価する課題とは言えないと考えられる。

このような既存の四項類推の問題点を解消するために、我々は二項未知類推課題を提案する。二項未知類推課題は、関係の関係性を用いたの推論を必要とする課題の一つである。課題内容としては、「男」と「女」単語と同じ関係を持つ単語の対を 1 つあげよ(前述の四項類推課題と対比すれば“男:女::?:?”)という形式の問いへの回答を求める。この課題では、各単語への類似性などだけでは、無数にありうる単語対(1 単語を答えるよりも潜在的な組み合わせが大きい)を絞り込むことが困難であるため、問いに含まれる 2 つの単語の関係を基に推論することが不可欠であると考えられる。したがって、類推能力、つまり関係の関係性に基づく推論能力を評価する課題と言える。

3. 単語埋め込みモデルで二項が未知の類推課題を解けるか

以下では二項未知類推課題を word2vec のベクトル空間を用いて回答した結果を報告する。ここでは word2vec モデルとして訓練済みの単語ベクトル¹を用いた。この単語ベクトルが構成するベクトル空間は、300 次元上の各点として 300 万の単語ベクトルを持つ。

Word2vec のベクトル空間を用いた二項未知類推課題の回答方法の一つの実装として、候補単語対の集合の内、問いの 2 単語の差のベクトルとのユークリッド距離が最も小さい差のベクトルを構成する単語対を類推結果として回答すると定義した。このように演算を定

表 1: Google テストセットの問題クラスと各クラスの問題例

単語クラス	問題例
首都と主要な国	Berlin : Germany :: Paris : France
首都と世界の国	London : England :: Rome : Italy
通貨	Japan : yen :: USA : dollar
街と州	Boston : Massachusetts :: Honolulu : Hawaii
男女	man : woman :: king : queen
形容詞と副詞	amazing : amazingly :: calm : calmly
対義語	acceptable : unacceptable :: aware : unaware
比較級	bad : worse :: big : bigger
最上級	bad : worst :: big : biggest
現在分詞	code : coding :: dance : dancing
国籍	France : French :: Germany : German
過去形	dancing : danced :: decreasing : decreased
複数形	banana : bananas :: bird : birds
複数動詞	decrease : decreases :: describe : describes

義した理由は、単語ベクトル間の平行四辺形関係を適切に捉えるためである。ある図形が平行四辺形であるための条件は、対辺の長さとお角がそれぞれ等しいことである。 $u, v \in \mathbb{R}^n$ としたときのユークリッド距離 $\|u - v\|$ は二つのベクトルの差分を計算することでベクトル間の長さを比較し、差分の内積を計算することで角度を比較する。その点で、ユークリッド距離は複数ベクトルが平行四辺形をなしているかどうかを確認するのに適している。テストセットとしては、単語埋め込みモデルの類推課題で通常用いられるテストセットである Google テストセット¹を用いた(表 1)。テストセット中の各単語クラスについて、その 1 つの単語対(e.g., 「男」-「女」)を二項未知類推形式の問いとし、その単語対と同じ単語クラスの他の単語対(e.g., 「王」-「女王」など正解の事例群)の差のベクトルと

¹ <https://code.google.com/archive/word2vec>

のユークリッド距離と、問いの単語対と不正解の単語対の集合 (e. g., 「東京」 - 「りんご」) から無作為抽出した 1000 単語対のユークリッド距離を比較した (表 2). ある単語クラスで、同クラスの他の単語対が問いの単語対に対して、不正解単語対より小さい距離を持つならば、単語埋め込みモデルのベクトルを使って、二項未知類推課題で正解できることを意味する.

この分析を行った結果、「男女」、「複数動詞」二つのカテゴリを除いた他のカテゴリでは、この回答方式により二項未知類推課題で正解できたケースは 0 であり、複数動詞カテゴリでも 1 つの対で正解できただけであった. また今回の実験では 1000 の単語対をランダムにサンプルしたが、回答方式として一つでも正解単語対より問いの単語対との距離が小さい不正解単語対があ

表 2: 分析したテストセットの各単語クラスに含まれる単語対の組み合わせ数のうち、不正解単語対の集合からランダムに抽出した 1000 単語対よりも、正解単語対の距離が小さい単語対の数(モデルの正しい回答数に相当).

単語クラス (単語対数)	正答数
首都と主要な国	0/22
首都と世界の国	0/115
通貨	0/29
街と州	0/66
男女	9/22
形容詞と副詞	0/31
対義語	0/28
比較級	0/36
最大級	0/33
現在分詞	0/32
国籍	0/40
過去形	0/39
複数形	0/36
複数動詞	1/29

った時点で不正解となるため、サンプル数を増やしても二項類推課題の精度が改善することはない. この結果は、word2vec のベクトル空間を用いて関係の間の関係性に基いて二項未知類推課題に高い精度で回答することは困難であることを示唆する. 従って、従来 word2vec は一項未知の類推の上で一部の類推課題を解けるとされてきたが (Chen et al., 2016; Lu et al., 2019), 関係の関係性を捉える、という意味で課題を解けているのかどうかは疑問が残る.

4. おわりに

本稿では、既存の一項が未知の四項類推課題では人の類推能力をテストできていない可能性を指摘し、それに代わる課題として二項が未知の四項類推課題を提案した. この課題を用いて word2vec の性能をテストし、word2vec は二項が未知の類推課題を高い精度では解けないことを示した.

また、今回 word2vec を使った場合二項未知類推課題の回答の精度が低いことを示した. 二項未知類推課題に回答可能なモデルを作るために以下のような方策が考えられる. (a) 例えば男女関係のような、ある関係にとって重要となる次元を選定し、その次元を重み付けるなどの手法で関係毎の部分空間を作り、その上で類推を行う. (b) neural latent relational analysis (Washio and Kato, 2018) や pair2vec (Joshi et al., 2018) など、単語対のベクトル表現を生成する手法を用いる. このようにして二項未知類推課題に高精度で回答できるモデルを構築することで、類推を行う能力にとってどのような性質が必要かつ十分となるのかを明らかにすることが期待できる.

参考文献

- [1] Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 238-247).
- [2] Chen, D., Peterson, J. C., & Griffiths, T. L. (2017). Evaluating vector-space models of analogy. *arXiv preprint arXiv:1705.04416*.
- [3] Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41(1), 1-63.
- [4] Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science*, 7(2), 155-170.
- [5] Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cognitive Science*, 34(5), 752-775.
- [6] Holyoak, K. J., & Thagard, P., 鈴木宏昭, & 河原哲雄. (1995). 『アナロジーの力』. 新曜社, 東京.
- [7] Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110(2), 220-264.
- [8] Joshi, M., Choi, E., Levy, O., Weld, D. S., & Zettlemoyer, L. (2018). pair2vec: Compositional Word-Pair Embeddings for Cross-Sentence Inference. *ArXiv:1810.08854 [Cs]*.
- [9] Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences*, 116(10), 4176-4181.
- [10] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- [11] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and

their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

- [12] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- [13] Rogers, A., Drozd, A., & Li, B. (2017, August). The (too many) problems of analogical reasoning with word vectors. In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)* (pp. 135-148).
- [14] Washio, K., & Kato, T. (2018). Neural Latent Relational Analysis to Capture Lexical Semantic Relations in a Vector Space. *ArXiv:1809.03401 [Cs]*.