

# 知識の階層図による理解度評価を目的とした Word2vec の試用

## Trial of Using Word2vec for Assessment of Understanding by Knowledge Hierarchy

大崎理乃<sup>1</sup> 柴田淳司<sup>1</sup>

Ayano OHSAKI<sup>1</sup>, Atsushi SHIBATA<sup>1</sup>

<sup>1</sup> 産業技術大学院大学 産業技術研究科

<sup>1</sup> Graduate School of Industrial Technology, Advanced Institute of Industrial Technology

**Abstract:** The purpose of this study is development of an assessment system for learning as knowledge-creation. It is an important viewpoint to assess how people use knowledge for a learning environment design and re-design. Furthermore, real-time analysis and time series analysis are effectiveness for facilitation by a teacher in a classroom. The first author suggested using knowledge hierarchy for assessment collaborative learning in the previous study. However, it needs time to collate utterances with knowledge. Hence, we proposed an analysis system using Natural Language Processing. In this study, we calculated similarities between an utterance and knowledge in collaborative learning by using Word2vec. As a result, we confirmed that the differences between an external model and a knowledge hierarchy model.

### 1. はじめに

近年の学習研究では、学習プロセスが重要な評価対象とされている。例えば、授業改善の視点として「どのように学ぶか」が明示されている[1]ほか、学習科学領域の研究手法では、学習プロセスでの設計原則の実現を確認することが求められている[2]。

知識創造の観点では、学習プロセスにおける学習者の知識利用と知識創造の状況を把握することが、指導や学習環境のデザインに必要とされる。そこで第1著者は、知識の広さと深さを可視化する尺度として図1のような知識の階層図(以下、知識階層図)を開発し、学習プロセスや学習成果物の分析を通じた知識利用状況確認の有用性を報告してきた [3-4]。これらの成果から、知識階層図を用いた学習プロセスの時系列分析やリアルタイム分析は、授業中の学習者状況を教師が把握することや指導に有用であると考えられる。

しかし先行研究[3]では、分析者による分析のばらつきや手動での分析に時間かかるといった課題が残された。この点は、学習プロセスの分析結果を指導に利用する上で大きな問題となる。そこで本研究では、学習者データと階層図で表現された知識の類似度を算出する方法を検討した。

\*連絡先：産業技術大学院大学産業技術研究科

〒140-0011 東京都品川区東大井 1-10-40

E-mail: ohsaki-ayano@aait.ac.jp

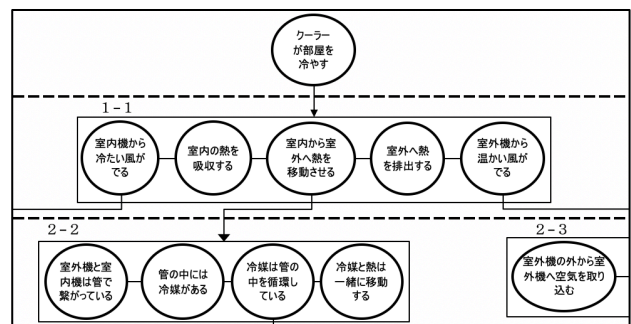


図1: 知識階層図の例[3].

### 2. 提案

本研究では、学習者の理解状況を評価するために、学習者の発話を自然言語処理によって知識の階層図上へマッピングすることで、学習プロセスを分析するシステムの構築を目指す。学習プロセスでの知識利用の変化や頻度を把握することは、学習者の理解度を評価することにつながると著者らは考えた。

そこで、本稿では Word2vec による話者の発話と階層図における知識の類似度を検討対象とする。Word2vec は自然言語処理の一種であり、Bag of Words と共に単語間のマッチングに多く使われている。Bag of Words は文書の特徴語抽出を目的として用いられる一方、Word2vec は語の意味を計算可能とする方法として知られている。Word2vec による意味

の数値化は様々な分野で注目されており，Zhu et al.[5]は生物医学用語間の意味と類似度を数値化したデータベースの開発に Word2vec を使用している．本研究では，発話や記述などの学習者データ文と，知識階層図における文の類似度を計算することを目指しているため，語の意味を計算可能な Word2vec を採用することとした．

### 3. 実験

Word2vec による分析方法の探索的検討として，知識階層図モデルと外部モデル，異なる二つのモデルを使用した実験を行った．知識階層図モデルは，分析対象データに対応した知識階層図を利用して，著者らの生成したモデルである．モデルの生成には，gensim を利用し，設定は size=200, min\_count=1, window=15 とした．外部モデルは Wikipedia のデータを元に作製されたモデル[6]を採用した．また，実験には gensim の n\_similarity を利用し，各文の類似度を算出した．

実験に利用した学習データは，教育学系ゼミの大学院生と大学生を対象としたワークショップにて参加者の同意を得て収集した．活動は「クーラーが部屋を冷やし続ける仕組み」の説明資料作りであった．参加者は事前のグループ活動で，資料 A「熱の性質」，資料 B「物質の状態変化と熱」，資料 C「圧力と温度」の 3 種類についてそれぞれ異なるグループで学習した後，各資料を担当したメンバーが一人ずつ集まったグループで説明資料を作製した．

本分析では，無作為に抽出した 1 グループの発話から「熱が移動するんよ．そしたらその冷媒って冷める」(以下，発話)という文と，当該文に関連があると著者らが判断した知識階層図上の文「冷媒から空気へ熱が移動する」(以下，知識 1)，「空気の温度が上がる」(以下，知識 2)，「冷やされた冷媒が室外機から出る」(以下，知識 3)を使用した．

### 4. 結果

実験の結果を表 1 に示す．まず，採用した外部モデルには「冷媒」「空気」といった教材の重要単語が含まれておらず，本研究で目指す意味の数値化に対して単純な適用はできないことが確認された．例えば知識 1 は，発話と同じ「熱」「移動」のみが分析に使用可能な単語となり，発話に対して極めて高い類似度が算出された．また，知識 2 では，分析に使用可能な単語が「上」だけとなり，発話に対する類似度は極めて低く算出された．

一方で知識階層図モデルは，周辺意味を加える必

要があるものの，分析に利用可能であることが示唆された．例えば，文字的に近い発話対知識 1 は類似度が高く，一部文字的に近い発話対知識 3 は一定の類似度が算出された．しかし，意味的に近いものの文字的には遠い発話対知識 2 は類似度が低く，モデルの調整が必要と考えられる．

表 1: 発話と階層図上知識の類似度計算結果.

	知識階層図 モデル	外部モデル
発話 対 知識 1	0.922	1.000
発話 対 知識 2	0.371	-0.038
発話 対 知識 3	0.609	0.213

今回の実験では，モデルの違いによる分析結果の差異に焦点を絞り分析方法を検討した．今後，複合語や品詞なども考慮して生成されたモデルの利用や，前処理を含めた分析の手続きを検討していく予定である．

### 謝辞

本研究の一部は，科研費 JSPS 科研費 18K13238 の助成を受けた．

### 参考文献

- [1] 文部科学省：新学習指導要領，[http://www.mext.go.jp/a\\_menu/shotou/newcs/1383986.htm](http://www.mext.go.jp/a_menu/shotou/newcs/1383986.htm) (2017) (2020.2.7.参照)
- [2] Sandoval, W.: Conjecture mapping: An approach to systematic educational design research, *Journal of the Learning Science*, Vol. 23, No. 4, pp. 18-36 (2014)
- [3] 大崎理乃，山田雅之：協調学習における知識利用状況の機能機構階層図による可視化，*日本教育工学会論文誌*，Vol.42, No. Suppl, pp.85-88 (2018)
- [4] 大崎理乃，千代西尾祐司，村上隆正：中学校理科での協調学習を通じた説明資料作成活動における高自信群成果物の分析，*日本教育工学会，教育工学会研究報告集*，JSET18-4，pp. 71-74, (2018)
- [5] Zhu, Y., Yan, E. & Wang, F.: Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec, *BMC Med Inform Decis Mak*, Vol. 17, Paper No. 95 (2017)  
<https://doi.org/10.1186/s12911-017-0498-1>
- [6] Kyubyong Park: wordvectors , <https://github.com/Kyubyong/wordvectors> (2016) (2020.2.7.参照)