Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Method for Extracting Unique Meanings of Synonyms Using Word2Vec for Vocabulary Learning

Yuto KAKUTA[a*], Daiki MAEDA[a] & Tomoko KOJIRI[b]
[a] *Graduate School of Science and Engineering, Kansai University, Japan*
[b] *Faculty of Engineering Science, Kansai University, Japan*
*k146565@kansai-u.ac.jp

Abstract: In this study, we propose a method for extracting word unique meanings of synonyms using Word2Vec for the purpose of supporting vocabulary learning. In conventional vocabulary learning, it was difficult for the system to teach the meaning differences between synonyms. This study attempts to propose a method for extracting words that represent unique meaning of each synonym pair using Word2Vec. Our method derives a vector of a unique meaning of the target word by subtracting the common vector of the two words from the vector of the target word. In the verification experiment, the proposed method was able to extract unique meaning of Japanese synonym pairs that contain one same Chinese character at a maximum rate of 64%.
Keywords: vocabulary learning, unique meaning, Word2Vec

1. Introduction

To use synonyms appropriately, learners must understand not only each word's meaning but also subtle distinctions between them. For example, "study" and "learn" both mean acquiring knowledge or skill, but "study" emphasizes progress or process, whereas "learn" emphasizes result or achievement. If this difference is unclear, one may be used where the other one is more appropriate. Therefore, this study aims to support learners in mastering the appropriate use of synonym pairs according to the given context.

Conventional dictionaries describe word meanings in detail but rarely explain how similar words differ. Synonym dictionaries list related terms, but do not describe unique meaning of all pairs. The unique meaning is different according to the synonym pair. For instance, the unique meaning of "study" from "investigate" is to deepen knowledge or understanding. Therefore, it is difficult to prepare unique meaning of all synonym pairs. Automatically extracting word's unique meaning of synonyms are useful.

Existing vocabulary learning systems, such as (Toniolo et al., 2022), focused on teaching meaning of individual words rather than differences between them. Wang proposed the system that detects inappropriate words in learner-generated sentences and suggests better alternatives (Wang, 2023), but it does not explicitly show how synonyms differ.

Machine learning models such as Word2Vec represent words as multidimensional vectors. It is known that addition and subtraction of vectors correspond to addition and subtraction of meanings. Based on this, we propose a method for extracting words using arithmetic operations of word vectors. Our method derives a vector of a unique meaning of the target word by subtracting the common vector of the two words from the vector of the target word. Our method derives a vector of a unique meaning of the target word by subtracting the common vector of the two words from the vector of the target word.

2. Word2Vec

Word2Vec is a model that represents the meaning features of words as vectors in a multidimensional continuous space (Mikolov et al., 2013). The words with similar meanings are learned to be close to each other in the space. By taking the cosine similarity of vocabulary vectors, it is possible to express the similarity between words. It is also possible to perform arithmetic operations on word vectors. For example, it is known that the result of "king" - "male" + 'female' is a vector similar to the vector of "queen".

3. Extraction Method of Unique Meaning Words of Synonym Using Word2Vec

We assume that vectors that represent unique meaning can be derived by adding word-unique components to that word. This word-unique components can be derived by decreasing common components from the word. Based on these assumption, we propose a method for

extracting words that can represent unique meanings as Equations (1) and (2). Figure 1 shows a conceptual diagram of the vectors represented by equations (1) and (2).

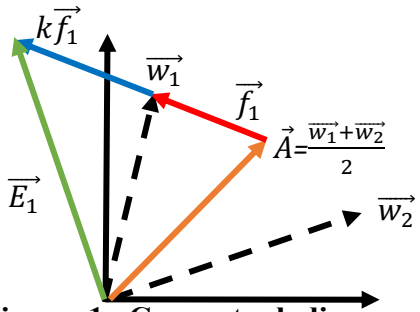

Figure 1. Conceptual diagram of equations (1) and (2)

$\overrightarrow{w_1}$and $\overrightarrow{w_2}$ are word vectors of the synonym pair. $\overrightarrow{f_1}$ is a unique component of w_1, and $\overrightarrow{E_1}$ indicates the word vectors that represent unique meaning w_1. k is a constant that emphasizes $\overrightarrow{f_1}$. $\overrightarrow{f_1}$ is derived by subtracting the average vector of synonyms $\frac{\overrightarrow{w_1}+\overrightarrow{w_2}}{2}$ from $\overrightarrow{w_1}$.

$$\overrightarrow{E_1} = \overrightarrow{w_1} + k\overrightarrow{f_1} \quad (1)$$

$$\overrightarrow{f_1}=\{\overrightarrow{w_1} - (\frac{\overrightarrow{w_1} + \overrightarrow{w_2}}{2})\} \quad (2)$$

4. Evaluation Experiment

We have conducted an experiment to evaluate the effectiveness of the proposed method. The experiment targeted Japanese synonyms. The Wikipedia corpus in Japanese was used for Word2Vec training. The gensim library was employed, with the vector dimension set to 300 and the number of epochs set to 9.

4.1 Verification Experiment1: Average Vectors

We have conducted an experiment to verify whether the average vector $\frac{\overrightarrow{w_1}+\overrightarrow{w_2}}{2}$ represents the common meaning of two synonyms. We used 45 synonym pairs from the Kotobank Digital Daijisen Dictionary. Table 1 shows the 45 synonym pairs used in this experiment. Common synonyms of these synonym pairs are acquired as words that represent common meanings (which we call answer words). We have evaluated whether the words whose cosine similarity with average vector are higher than 0.6 contain at least one answer word.

Table 1. 45 pairs of synonyms used in experiment1

ID	Pairs of Synonyms	ID	Pairs of Synonyms	ID	Pairs of Synonyms	ID	Pairs of Synonyms
1	感動・感心	13	収集・蓄積	25	経歴・履歴	37	採用・採択
2	推定・推測	14	連絡・接続	26	迅速・敏捷	38	完了・達成
3	出版・刊行	15	器具・道具	27	信頼・信用	39	真実・事実
4	購入・購買	16	辞書・辞典	28	探索・調査	40	役割・責任
5	解答・回答	17	危険・危惧	29	堅実・地道	41	定義・意味
6	準備・用意	18	心配・不安	30	記入・記載	42	主義・理念
7	形態・形状	19	感触・感覚	31	落胆・失望	43	知能・頭脳
8	早急・緊急	20	重要・大事	32	実行・実施	44	注意・配慮
9	承認・承諾	21	企画・計画	33	残念・遺憾	45	結果・結末
10	意図・目的	22	綿密・厳密	34	経費・費用		
11	提示・表示	23	適切・適当	35	案外・理解		
12	避難・逃避	24	業務・作業	36	禁止・規則		

The result is shown in Table 1 with the ID underlined. 26 underlined pairs successfully contained at least one answer word met this criterion. In addition, for pairs that have one identical kanji, 20 out of 26 pairs contain at least one answer word. This result indicates the assumption that the proposed method is effective particularly for pairs that share the identical character.

4.2 Verification Experiment2: Average Vector of Synonyms with One Identical Character

We have conducted an experiment to verify whether the average vector represents the common meaning of two synonyms that have the identical character. 31 new synonym pairs containing one identical character are selected as shown in Table 2. The method of the experiment is the same as experiment 1. The result is shown is Table 2 with ID underlined. 28 underlined pairs (90.3%) conained at least one answer word. This result supports our assumption that the average vector of synonyms with one identical word is able to capture common meaning words.

Table 2. 31 pairs of synonyms used in experiment2

ID	Pairs of Synonyms	ID	Pairs of Synonyms	ID	Pairs of Synonyms	ID	Pairs of Synonyms
1	発刊・刊行	9	使用・利用	17	輸送・運送	25	維持・保持
2	損失・損害	10	模写・描写	18	問題・疑問	26	目的・目標
3	法律・法令	11	団結・結束	19	発展・伸展	27	描画・描出
4	類似・酷似	12	停止・中止	20	調査・精査	28	回復・復元
5	推定・推察	13	調子・様子	21	活用・利用	29	戦闘・戦争
6	順序・順番	14	理論・論理	22	要件・条件	30	意識・認識
7	勉強・勉学	15	隠蔽・隠匿	23	管理・監理	31	悲惨・惨劇
8	欲求・欲望	16	感動・感慨	24	合致・一致		

4.3 Verification Experiment3: Vector Representing Unique Meaning

We have conducted an experiment to verify whether the vector of unique meaning $\overrightarrow{E_1}$represents the unique meaning of words. 43 synonym pairs with one identical character were prepared: for 18 pairs both words had unique meaning words (Present in Both Words) and for 25 pairs only one word did (Present in Only One Word). The answer words are synonyms of the target words excluding the common synonyms. We set k=1/2,1,5,10,50 and examined whether the top 15 words with the highest cosine similarity to the vector of unique meaning contained answer words.

Table 3. Percentage of words containing unique meaning words among words surrounding unique emphasis vectors

k	1/2	1	5	10	50
Present in Only One Word	20%	20%	16%	8%	5%
Present in Both Words	64%	42%	39%	36%	28%

Table 3 shows the results. The vectors derived with k=1/2 for pairs whose unique words are presented in both words were the best. We need to examine what kind of synonym pairs are most suitable for our proposed method and which k is appropriate as our future work.

5. Conclusion

In this study, we proposed a method for deriving the unique meaning words of a given pair of synonyms using vocabulary vectors represented by Word2Vec. The results of the verification experiment showed that the average vector could express a common meaning word for synonyms containing one identical character. However, the unique meaning words are derived for only 64% of synonym pairs. For future work, we need to find synonym pairs that work well with our proposed method. In addition, we need to propose the method for using these results in vocabulary learning.

References

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space (No. arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Wang, C., Mao, S., Ge, T., Wu, W., Wang, X., Xia, Y., Tien, J., & Zhao, D. (2023). Smart Word Suggestions for Writing Assistance (No. arXiv:2305.09975). arXiv. https://doi.org/10.48550/arXiv.2305.09975

Weerasinghe, M., Biener, V., Grubert, J., Quigley, A. J., Toniolo, A., Pucihar, K. Č., & Kljun, M. (2022). VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations (No. arXiv:2207.00896). arXiv. https://doi.org/10.48550/arXiv.2207.00896