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

Educational Cone Model in Embedding Vector Spaces

Yo EHARA[a]
[a]*Faculty of Education, Tokyo Gakugei University, Japan*
*ehara@u-gakugei.ac.jp

Abstract: Human-annotated datasets with explicit difficulty ratings are essential in intelligent educational systems. Although embedding vector spaces are widely used to represent semantic closeness and are promising for analyzing text difficulty, the abundance of embedding methods creates a challenge in selecting the most suitable method. This study proposes the Educational Cone Model, which is a geometric framework based on the assumption that easier texts are less diverse (focusing on fundamental concepts), whereas harder texts are more diverse. This assumption leads to a cone-shaped distribution in the embedding space regardless of the embedding method used. The model frames the evaluation of embeddings as an optimization problem with the aim of detecting structured difficulty-based patterns. By designing specific loss functions, efficient closed-form solutions are derived that avoid costly computation. Empirical tests on real-world datasets validated the model's effectiveness and speed in identifying the embedding spaces that are best aligned with difficulty-annotated educational texts.

Keywords: Embedding, difficulty, vector spaces

1. Introduction

Datasets that are annotated with difficulty levels by educators ("difficulty-annotated educational datasets") are essential for developing educational support systems (Arase et al., 2022; Hendrycks et al., 2020). Although embedding spaces are widely used to represent semantic similarity and are promising for analyzing such datasets, the abundance of embedding methods (Muennighoff et al., 2022) poses a challenge in selecting suitable methods.

To address this, we propose the Educational Cone Model, which assumes that easier items covering fundamental concepts exhibit lower diversity, while more difficult items are more diverse. This results in a cone-like structure in the embedding space, independent of specific methods. This intuition aligns with the findings of vocabulary acquisition (e.g., Zipf's law), Piaget's developmental stages (Piaget, 1952), and Bloom's taxonomy.

We mathematically show that evaluating alignment with this model is reduced to solving an optimization problem that identifies a "difficulty direction" in the embedding space. By designing appropriate loss functions, we derive closed-form solutions, avoiding computationally expensive operations such as centroid comparisons.

Empirical evaluations with recent sentence embeddings confirm that the proposed model enables efficient selection of embedding methods that are well-aligned with difficulty-annotated datasets.

The contributions of this study are summarized as follows: 1) We propose a geometric model that reflects the assumption that easier items are less diverse in the embedding space. 2) We show that the model can identify a difficulty direction via optimization. 3) We formulate this as a closed-form optimization problem. 4) We demonstrate that this solution requires only mean vector differences between difficulty levels. 5) We empirically validate the effectiveness of the model on real datasets using recent sentence embeddings.

2. Formulation

2.1 Notation

Let $\{x_1, . . ., x_N\}$ denote a set of *N* embedding vectors, where $\boldsymbol{x_i}$ is a *D*-dimensional vector. We assume that all embedding vectors are normalized; that is, $||\mathbf{x}|| = 1$, where $||\mathbf{x}||$ denotes the Euclidean norm of the vector. The proposed method can be applied to both word and sentence embeddings if the aforementioned conditions are satisfied. For simplicity, we refer to both words and sentences as items throughout this paper. We introduce a *D*-dimensional vector $\mathbf{w} \in \mathbf{R}^D$ to represent the direction in the embedding space, with the aim of determining the coordinates of **w**.

2.2 Educational Cone Model and Difficulty Direction Search Problem

We first consider the Educational Cone Model and show that, under simple assumptions, it aligns with the problem of searching for the difficulty direction in the embedding space. The Educational Cone Model assumes that simpler items exhibit lower diversity and more difficult items exhibit higher diversity. We interpret the magnitude of diversity in terms of the spatial spread within the embedding space. Assuming that simpler items exhibit lower diversity, their embedding vectors have a smaller spread in the embedding space. By further refining the notion that simpler items exhibit lower diversity, we assume the existence of the "simplest item." Embedding vector spaces are typically structured such that semantically similar items are positioned closer together. Hence, in the embedding vector space, the simplest item is assumed to reside in the least-spread-out region, represented by point $\boldsymbol{e}$. If we consider difficulty as a component of "semantic similarity," items closer to $\boldsymbol{e}$ should be simpler, whereas those farther away should be more difficult. In the Educational Cone Model, suppose that $\boldsymbol{x_i}$ is simpler than $\boldsymbol{x_j}$. Based on the above discussion, $\boldsymbol{x_i}$ is closer to the "simplest item" $\boldsymbol{e}$ than $\boldsymbol{x_j}$. By measuring the distance using the Euclidean distance and assuming that all x are normalized to $||\mathbf{x}|| = 1$, we obtain the following transformations: $||\boldsymbol{x_i} - \boldsymbol{e}|| < ||\boldsymbol{x_j} - \boldsymbol{e}|| \Leftrightarrow \boldsymbol{w}^\top \boldsymbol{x_i} < \boldsymbol{w}^\top \boldsymbol{x_j}$. Here, we define $\boldsymbol{w}$ as $-\boldsymbol{e}$. $\boldsymbol{w}$ represents a direction in the embedding vector space. The expression $\boldsymbol{w}^T \boldsymbol{x_i}$ implies that items can be arranged in order along this direction. Consequently, the direction $\boldsymbol{w}$ indicates that moving in this direction within the embedding vector space corresponds to increasing difficulty.

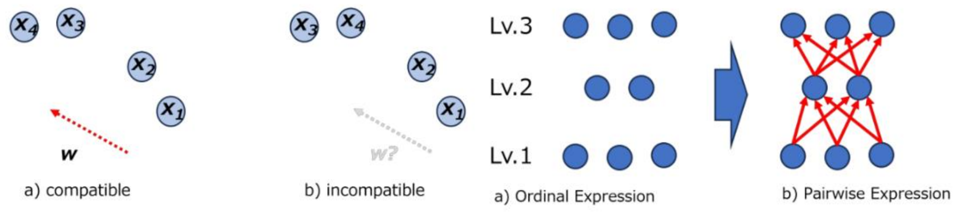

Figure 1. Left: Overview of the proposed method: a) Considering $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3},$ and $\boldsymbol{x_4}$ as two-dimensional (2D) word/sentence embeddings. That is, $\boldsymbol{x_1}$ is annotated as simpler than $\boldsymbol{x_2}$ which, in turn, is simpler than $\boldsymbol{x_3}$, etc. If listing points along direction **w** aligns with the annotation, the embedding vector set is defined as "compatible" with the annotation. b) In this case, no direction in the 2D space orders $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3},$ and $\boldsymbol{x_4}$ in the annotated order, so the embedding is defined as "incompatible." Right: Conversion of difficulty annotations into pairwise constraints.

We provide an intuitive interpretation of the difficulty direction $\boldsymbol{w}$. Simply put, the difficulty direction $\boldsymbol{w}$ represents the direction in which all items in the dataset (annotation set) appear, arranged in order of difficulty. Although determining an ideal direction is preferable, it is often unrealistic. To address this, we allow slight deviations in the order of difficulty. To this end, we first introduce the concept of "compatibility" in Figure 1 (left). In practice, difficulty annotations are often provided in ordinal levels, rather than as direct pairwise relationships. For example, a question might be annotated as high-school or university level rather than being directly compared to another question (Figure 1, right). As shown in (a), eight items are annotated using three levels: Levels 1, 2, and 3. These levels, abbreviated as “Lv,” indicate increasing difficulty with higher numbers. Levels 1, 2, and 3 have three, two, and three items, respectively. The entire ordinal structure can be converted into a directed graph as shown in (b). The directed edges signify pairwise relationships, and an edge from node *i* to node *j* indicates that *i* is easier than *j*. In this manner, without loss of generality, any finite set of ordinal annotations can be converted into a mathematically equivalent directed graph.

As illustrated in Figure 1, the difficulty annotations can be converted into a set of pairwise constraints. To ensure generality, we define a set of pairwise comparison constraints, in which each constraint indicates that one embedding vector is annotated more easily than the other. We define the set of pairwise order constraints as C = $\{(i_1, j_1), . . ., (i_K, j_K)\}$, where K denotes the number of constraints. The k-th pair (i_k, j_k) represents a single constraint.
Here, $i_k \in \{1, . . ., N\}$ and $j_k \in \{1, . . ., N\}$ denote indices of N embedding vectors, each of which represents the meaning of each item. The annotation (i_k, j_k) indicates that x_{ik} is easier than x_{jk}. For simplicity, we omit the subscript k and denote the easier vector as x_i and the more difficult vector as x_j. We then project these vectors onto the direction defined by $\boldsymbol{w}$ to model the ordering. Let θ_i and θ_j denote the angle between x_i and the direction $\boldsymbol{w}$ and the angle between x_j and $\boldsymbol{w}$, respectively. Our goal is to adjust $\boldsymbol{w}$ such that the pairwise constraints are satisfied: $||\boldsymbol{x}_i|| \cos \theta_i < ||\boldsymbol{x}_j|| \cos \theta_j \Leftrightarrow ||\boldsymbol{w}||||\boldsymbol{x}_i|| \cos \theta_i < ||\boldsymbol{w}||||\boldsymbol{x}_j|| \cos \theta_j \Leftrightarrow \boldsymbol{w}^\top\boldsymbol{x}_i < \boldsymbol{w}^\top\boldsymbol{x}_j \Leftrightarrow \boldsymbol{w}^\top (\boldsymbol{x}_i - \boldsymbol{x}_j) < 0$. If K pairwise constraints exist, they must all hold simultaneously. However, for practical datasets with many items, there may not exist a $\boldsymbol{w}$ that meets all K constraints. To this end, we introduce a slack variable ξ_k to rewrite the inequality constraint into an equality constraint: $\mathbf{w}\top (\boldsymbol{x}_{ik} - \boldsymbol{x}_{jk}) + \xi_k = 0$. Intuitively, ξ_k represents the degree to which the constraint is maintained. A larger value of ξ_k means that a greater margin is preserved, whereas a smaller value indicates a minimal margin enforcement. While $\xi_k \geq 0$, we remove this constraint to relax the problem:

$$\text{maximize}_{\xi,\boldsymbol{w}} \sum_{k=1}^{K} \xi_k \quad \text{s.t.} \quad \forall k \in \{1, \dots, K\}; \boldsymbol{w}\top(\boldsymbol{x}_{ik} - \boldsymbol{x}_{jk}) + \xi_k = 0, \left||\boldsymbol{w}|\right|^2 = 1. \quad (1)$$

In Equation (1), we also impose a norm constraint $||\boldsymbol{w}||^2=1$ to obtain a fixed-form solution. First, by noting that $\xi_k = -\boldsymbol{w}\top(\boldsymbol{x}_{ik} - \boldsymbol{x}_{jk})$ and introducing a Lagrange multiplier λ to enforce the equality constraint $1 - ||\mathbf{w}||^2$, we obtain the following unconstrained optimization problem:

$$\text{maximize}_{\boldsymbol{w}} \sum_{k=1}^{K} -\boldsymbol{w}\top(\boldsymbol{x}_{ik} - \boldsymbol{x}_{jk}) + \lambda\left(1 - \left||\boldsymbol{w}|\right|^2\right). \quad (2)$$

Differentiating Equation (2) with respect to $\boldsymbol{w}$, we obtain $\sum_{k=1}^{K} -(\boldsymbol{x}_{ik} - \boldsymbol{x}_{jk}) - 2\lambda\boldsymbol{w} = 0$, from which we obtain $\boldsymbol{w} \propto \sum_{k=1}^{K}(\boldsymbol{x}_{jk} - \boldsymbol{x}_{ik})$. This confirms that the optimal direction is simply the mean of the difference vectors normalized to the unit length. This property enables us to determine the optimal vector efficiently without solving the optimization problem each time.

3. Experiments

3.1 Consistency Experiment Using Word Embeddings and Fine-Grained Annotations

Consistency experiments were conducted using word embeddings. The dataset used was the CEFR-J Vocabulary Profile, which contains manually annotated word difficulty levels based on the Common European Framework of Reference for Languages (CEFR) (https://github.com/openlanguageprofiles/olp-en-cefrj?tab=readme-ov-file). FastText (Bhattacharjee, 2018) was used as the word embedding model. We employed the Support Vector Machine (SVM) as the baseline method. The regularization parameter C was tuned using the validation data, and the optimal value was selected from {0.1, 1.0, 10.0}. To verify the consistency, we followed a procedure that categorizes words into four levels of difficulty. A total of 100 words were randomly selected from the second-easiest category. For each word, pairwise constraints were formulated by treating pairs in the easiest category as training data and pairs in the third-easiest category as test data. Thus, the proposed convex optimization problem was solved. Both the proposed method and SVM achieved 100% accuracy.

In this approach, each word can be regarded as containing 100 different subset datasets. Another dataset, SVL, provides a more fine-grained 12-level annotation of word difficulty (https://eow.alc.co.jp/svl_level12.html). Using these 100 datasets, we assessed the annotation consistency based on the objective function value of the proposed convex optimization problem. As the annotation consistency is measured with respect to the simplest word category, words with higher difficulty levels in the 12-level SVL dataset were expected to exhibit greater consistency. Spearman's rank correlation analysis confirmed this expectation and yielded a statistically significant correlation ($p < 0.01$). These results demonstrate that the proposed method effectively detects consistency in word annotations.

Table 1. *Compatibility scores calculated using training data. "Model" represents each individual model, and each column corresponds to a pair of CEFR levels. The compatibility score indicates the degree of fit between the level annotations and each embedding. The first two models are 384-dimensional and the latter two models are 1,024-dimensional.*

Model	(A1, A2)	(A1, B1)	(A1,B2)	(A2,B1)	(A2,B2)	(B1,B2)
all-MiniLM-L6-v2	0.3227	0.5741	0.9308	0.4183	0.8599	0.7986
multilingual -e5-small	0.0835	0.1456	0.2301	0.0835	0.1961	0.1782
bge-m3	0.1354	0.2317	0.3930	0.1410	0.3516	0.3336
multilingual -e5-large	0.0826	0.1435	0.2256	0.0844	0.1952	0.1756

3.2 Consistency Experiment Using Sentence Embeddings

The proposed method is applicable to both word and sentence embeddings. To examine its utility further, we evaluated its ability to predict annotation inconsistencies in sentence embeddings. For this experiment, we used the CEFR-SP dataset (Arase et al., 2022), which consists of English sentences annotated with four difficulty levels by two annotators. Following the same approach as that in the word-level experiment, we applied the proposed convex optimization problem to the dataset of one annotator. Sentence embeddings were generated using 'multilingual-e5-small' 3. Owing to the norm constraint in the proposed convex optimization, the solution vector lengths were approximately 1.0, enabling a direct comparison of the objective function values across sentences. As the value of ξ can be interpreted as a margin, a higher value indicates greater consistency. Next, we compared the results with those of the second annotator and calculated the Spearman's rank correlation coefficient between the annotation agreement and objective function value. The correlation coefficient was 0.262, indicating a weak correlation but without statistical significance. This suggests that although the proposed method captures annotation inconsistencies to a certain extent, further investigation is required.

Using the CEFR-SP dataset, we conducted experiments to identify the most compatible sentence embedding model. Each difficulty level pair was treated as a separate dataset and the compatibility score was computed using the proposed method. Table 1 presents the compatibility scores obtained from the sentence embedding experiments. Each column corresponds to a CEFR difficulty level pair, and Table 1 summarizes the embedding models used. Higher scores indicate a better fit between the embedding space.

3.3 Prediction Experiment Using SVM

Finally, we conducted a prediction experiment using SVM with sentence embeddings. We utilized (A1, B1) as the training datasets and (B1, B2) as the test datasets. The four embedding models summarized in Table 1 achieved prediction accuracies of 0.26, 0.23, 0.63, and 0.37, respectively. Notably, the first two embeddings had 384 dimensions, whereas the latter two had 1,024 dimensions. We observed that when the dimensionality was the same, embeddings with higher compatibility scores tended to yield superior predictive performance. This finding suggests that, given the same dimensionality, the compatibility score can reliably estimate the predictive performance of SVM without requiring access to the test data. Consequently, our approach enables an efficient assessment of the embedding quality for text difficulty prediction, without the need for computationally expensive model training on each embedding.

4. Related Work

As an early study on the relationship between contextualized embedding vectors and difficulty, Ehara (2022) expressed the difficulty of word examples by the frequency of nearby examples in the embedding vector space. Recently, Ehara (2025) proposed a method for controlling the difficulty of educational items by combining linear interpolation of sentence embeddings with different difficulty levels in the embedding space and a technique called Inverse Embedding, which generates sentences from embedding vector coordinates. Several recent studies have investigated the interpretability of embedding vectors (Vasilyev et al., 2024; Li & Li, 2024; Chen et al., 2024). Vasilyev et al. (2024) examined multilingual embeddings using linear transformations under the assumption that embeddings are orthogonally aligned across languages. Li and Li (2024) analyzed the relationships between contextualized embeddings from large-scale language models and static embeddings such as sentence embeddings. Chen et al. (2024) proposed methods for constructing finer-grained embeddings for thematic sentence representations. Building on this research, our study focuses on the consistency of ordinal annotations and presents a method for quantifying this consistency.

5. Conclusion

This paper has introduced the Educational Cone Model, which posits that easier items exhibit lower diversity than more difficult items in embedding vector spaces owing to their inherent complexity. This principle aligns with the findings of previous studies on educational classifications. Leveraging this insight, we proposed a computationally efficient method for determining the directional orientation of educational items within an embedding space. As a convex optimization approach, our method guarantees convergence to a global optimum, making it robust to the initial conditions and free from approximation errors. This contrasts sharply with recent neural-network-based methods, which are typically sensitive to initialization and only achieve locally optimal solutions.

The empirical evaluation demonstrated a statistically significant correlation between our method and fine-grained annotation consistency in word difficulty datasets. We further conducted experiments using the proposed method with recent sentence embedding. The proposed objective function accurately predicted the performance of embeddings as predictive features without using test data. This result substantiates our method’s ability to capture the “cone” structure of learning material difficulty within an embedding space effectively.

Future research directions include extending the approach to a broader range of educational tasks, including STEM-related questions. Notably, the interpretability of our method allows the identification of subject-specific difficulty dimensions, such as those unique to physics or chemistry, by treating difficulty as a directional property within the embedding space.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 22K12287 and by JST, PRESTO Grant Number JPMJPR2363.

References

Arase, Y., Uchida, S., & Kajiwara, T. (2022). CEFR-based sentence difficulty annotation and assessment. arXiv:2210.11766. https://doi.org/10.48550/arXiv.2210.11766

Bhattacharjee, J. (2018). fastText quick start guide: Get started with Facebook's library for text representation and classification. Packt Publishing.

Chen, S., Zhang, H., Chen, T., Zhou, B., Yu, W., Yu, D., Peng, B., Wang, H., Roth, D., & Yu, D. (2024). Sub-sentence encoder: Contrastive learning of propositional semantic representations. In K. Duh, H. Gomez, & S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, Long Papers) (pp. 1596-1609). Association for Computational Linguistics.

Ehara, Y. (2022). An Intelligent Interactive Support System for Word Usage Learning in Second Languages. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds) Artificial Intelligence in Education. AIED 2022. Lecture Notes in Computer Science, vol 13355. Springer, Cham. https://doi.org/10.1007/978-3-031-11644-5_37

Ehara, Y. (2025). Generating Diverse Difficulty Examples in Embedding Vector Spaces via Inverse Embedding. In: Cristea, A.I., Walker, E., Lu, Y., Santos, O.C., Isotani, S. (eds) Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED. AIED 2025. Communications in Computer and Information Science, vol 2592. Springer, Cham. https://doi.org/10.1007/978-3-031-99267-4_9

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2020). Measuring massive multitask language understanding. arXiv:2009.03300. https://doi.org/10.48550/arXiv.2009.03300

Li, X., & Li, J. (2024). BeLLM: Backward dependency enhanced large language model for sentence embeddings. In K. Duh, H. Gomez, & S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, Long Papers) (pp. 792-804). Association for Computational Linguistics.

Muennighoff, N., Tazi, N., Magne, L., & Reimers, N. (2022). MTEB: Massive text embedding benchmark. arXiv:2210.07316. https://doi.org/10.48550/arXiv.2210.07316

Piaget, J. (1952). The origins of intelligence in children (M. Cook, Trans.). International Universities Press.

Vasilyev, O., Isono, F., & Bohannon, J. (2024). Linear cross-lingual mapping of sentence embeddings. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024 (pp. 8163-8171). Association for Computational Linguistics.