

# Human-in-the-loop Generation System for Analogical Explanations of Target Concepts

**Ryo Takahashi**

Astemo, Ltd.  
Chiyoda-ku, Tokyo, Japan  
reizouko02.1199@gmail.com

**Asuka Terai**

Future University Hakodate  
Hakodate, Hokkaido, Japan  
aterai@fun.ac.jp

## Abstract

This study designs a human-in-the-loop system to generate analogical explanations for target concept pairs. The system first extracts a base concept for analogy from a corpus based on the target pair and focusing feature selected by the user. Then, a four-term analogy is performed to construct a base concept pair, each consisting of the base concept and corresponding concept, that reflects the relationships between the target concepts. Using either the base concept or base pair, a prompt is generated and provided to a large language model, which then produces an analogical explanation. Experiments were conducted to evaluate the system. The findings suggest that the parameter must be adjusted by the user and that the highest evaluation scores are obtained when using prompts that include only the base concept. Thus, the proposed system can be used to generate useful analogies to improve understanding of input concepts.

## 1 Introduction

Effective communication relies on the ability to intuitively explain complex concepts. Analogies, which draw parallels between a known concept and an unfamiliar one, provide a powerful rhetorical and cognitive tool that facilitates comprehension (Resnick et al., 2016) and enhances persuasion for attitude change (McCroskey and Combs, 1969). By linking new information (target) to familiar knowledge (base), analogies help bridge cognitive gaps and make abstract or complex ideas more tangible (Gentner, 1983) (Gentner and Gentner, 1983) (Richland et al., 2007).

Despite these advantages, the automatic generation of analogical explanations remains challenging owing to the complexity of identifying suitable analogies and producing coherent, contextually appropriate explanations. Recent advances in natural

language processing and large language models (LLMs) have opened new avenues for automated explanation generation. However, existing LLMs lack transparency regarding the selection of analogies for targets, and the explanations provided may not be appropriate for users.

Thus, this study proposes a human-in-the-loop system that involves human participation in LLM-driven systems in key stages of the workflow to improve transparency and provide appropriate output for users. The system generates analogical explanations by extracting suitable base concepts and base pairs for the target pair.

## 2 Related Work

### 2.1 Analogy

Analogical reasoning has been widely studied in both cognitive science and natural language processing (NLP). For example, the structure-mapping theory proposed by Gentner (Gentner, 1983) posits that analogy involves aligning the relational structures between a target and a base domain. This theory emphasizes deep structural similarity over surface-level features and has been applied to help explain analogies and metaphors.

In the NLP field, word-embedding techniques, such as Word2Vec (Mikolov et al., 2013), demonstrate that certain analogical relationships can be captured through vector operations. For instance, the relationship in “king is to man as queen is to woman” (four-term analogy) can be modeled using vector arithmetic. Although these models are effective for simple analogies, they often struggle with abstract or complex relational reasoning.

In four-term analogies, the relationship between the two concepts plays a crucial role. Bouraoui et al. (Bouraoui et al., 2019) proposed a method to improve relation induction by leveraging large corpora such as Wikipedia. They extracted sentences in which word pairs co-occurred and used them

---

<sup>1</sup>The first author conducted this study while they were students in Future University Hakodate.

as candidate relational templates. These sentences were then ranked using Bidirectional Encoder Representations from Transformers (BERT)(Devlin et al., 2019), and the top-ranked sentences were used to fine-tune models for relation induction tasks. Although this approach is promising for extracting rich relational features, it relies on the availability of multiple word pairs that share the same underlying relationship, which may limit its applicability in low-resource or open-domain settings.

## 2.2 Generation of Analogical Explanations

In studies that employ analogical explanations, Ishida et al.(Ishida et al., 1996) constructed a system based on a structural mapping theory(Gentner, 1983). This system generates explanatory sentences using metaphorical expressions for various terms within an intelligent manual system. It dynamically adapts the manner of explanation to the user’s knowledge. However, it has a limited knowledge base and can only generate explanatory sentences for predefined words.

Additionally, Chikamura et al. (Chikamura et al., 2016) proposed a method to generate explanations by collecting metaphorical sentences for a target from the web, extracting concepts for analogical explanations from the sentences, and evaluating them using the Kyodai case frame corpus(Kawahara and Kurohashi, 2016). In Japanese, case frames are particularly effective for evaluation because grammatical case markers explicitly encode semantic roles, enabling relations between concept pairs to be identified and compared more accurately. However, relying on analogies and metaphorical expressions found on the internet meant that it could only yield existing explanatory methods.

In this study, we assigned the target concept pair to a system, and the base concepts for analogy corresponding to one of the target concepts were extracted using a method that was improved by the previous method(Chikamura et al., 2016). This enabled the acquisition of flexible analogical representations. Subsequently, a four-term analogy was generated using existing embedding vector methods to create base concept pairs, each consisting of the base concept and the corresponding concept. Based on the method of Bouraoui et al.(Bouraoui et al., 2019), we extracted sentences from a large-scale corpus containing co-occurring word pairs and used this information to evaluate the base con-

cept pairs. An LLM then generated the analogical explanation using a prompt that included the base concept or pair selected by the user. The proposed system generated explanations interactively using appropriate analogies for users.

## 3 Generation System for Analogical Explanation

### 3.1 System Overview

The proposed system generates an analogical explanation for a target concept pair input by the user through four steps. An overview of this system is shown in Figure 1.

First, the user inputs a target concept (noun) pair (e.g., electric current and voltage). The system extracts features related to one of the target concepts and presents them to the user (e.g., electric current flows, measure voltage, etc.), and the user selects one as the focusing feature. The system extracts a base concept based on the focusing feature and target concept related to the feature. For example, if “electric current flows” is selected as the focusing feature related to one of the target concepts, the system focuses on the characteristics of the concept (“electric current”), especially the focusing feature (something that flows) when searching for suitable base concepts, such as “water.” The system then generates a four-term analogy using the target pairs and each base concept. For example, the system extracts concepts for  $x$  in the expression “electric current is to water as voltage is to  $x$ ”. These extracted base pairs, each consisting of the base concept and the corresponding concept, are evaluated based on the extent to which they share common features with the target pair. The base pairs with high scores are then presented to the user, who selects one to generate an analogical explanation. Finally, the LLM generates an explanation for the target pair using a prompt that includes the selected base concept or pair and the focusing feature.

### 3.2 Feature Extraction

The Wikipedia corpus is searched for sentences in which the target concepts co-occur, and dependency relationships are extracted from one concept of the pair using the dependency parser GiNZA. The extracted relationships are presented to the user as features of the target concept, and the user selects one as the focusing feature to generate explanations. Selecting a focusing feature enables

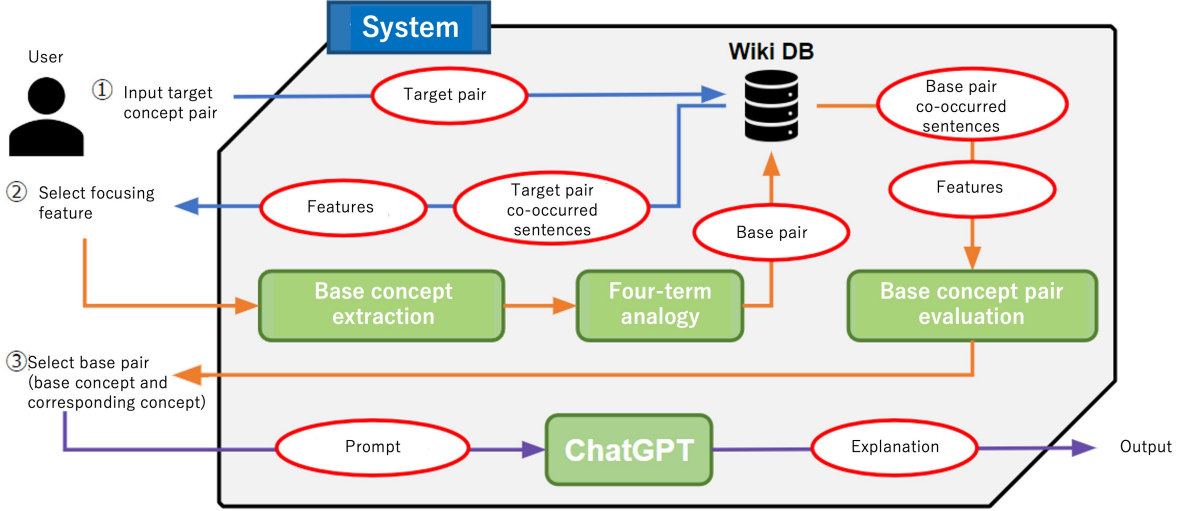


Figure 1: System overview.

the generation of appropriate explanations for the target pairs from the desired perspective of the user.

### 3.3 Extraction of Base Concepts for Analogy

#### 3.3.1 Relevance Score

Similar to Chikamura et al. (Chikamura et al. 2016), case frames were used to extract base concepts for the target concepts. Using the Kyoto University Case Frame Corpus (Kawahara and Kurohashi, 2016), usage examples were collected for each noun in the case frame, targeting nouns with frequencies of at least 100. The “NO (of)” frame was not used as it is generally used to express relationships between nouns. The relevance score between an input concept (word  $i$ ) and candidate concept (word  $j$ ) is defined as follows:

$$r_{ij} = \sum_{m \in M} \frac{f_m(i)}{\sum_{w \in W_m} f_m(w)} \frac{f_m(j)}{\sum_{w \in W} f_m(w)} \alpha_c, \quad (1)$$

where  $M$  denotes the set of all case frames (e.g. “\* GA NAGARERU (\* flow)”) shared by the words  $i$  and  $j$ ,  $W_m$  represents the set of all words having a given case frame and predication  $m$ ,  $f_m(w)$  is the frequency of the word  $w$  within the case frame  $m$ , and  $\alpha_c$  is the weighting applied to the case frame  $c$  corresponding to the focusing feature selected by the user, and the weighting is applied only when each frame  $m$  matches  $c$ .

#### 3.3.2 Concept Exclusion

To generate analogical explanations for the target concept using a different domain, concepts similar to the target concept are excluded from the

set of candidate base concepts. Thus, in DBpedia Japanese, headwords belonging to the same category as the target concept were removed from the candidate set.

Because explanations must be constructed using base concepts that users are already familiar with, the candidate base concepts were further filtered based on the Japanese Textbook Corpus Lexicon. This lexicon specifies the first stage of appearance of each word (concept). Words are classified into four stages based on the school level at which they first appear in Japanese textbooks: lower elementary, upper elementary, junior high, and high school. Consequently, words at later stages than the target concept were excluded from the candidate base concepts. If the target concept was not registered in the corpus, it was assumed to be more difficult than the words that appear in textbooks; thus, candidate words from all grade levels were considered.

### 3.4 Four-term Analogy

Using word embeddings generated by fastText (Bojanowski et al., 2016), a four-term analogy was produced based on the target concept pair and the base concept. Given a target pair consisting of words  $i_1$  and  $i_2$ , and the base concept ( $a_1$ ) extracted for the target concept ( $i_1$ ), the corresponding base concept ( $a_2$ ) is identified according to the following equation:

$$v(a_2) \approx v(a_1) - v(i_1) + v(i_2) \quad (2)$$

where  $v(*)$  denotes the embedding vector of each word. Words with embedding vectors similar to

$\hat{v}(a_2) = v(a_1) - v(i_1) + v(i_2)$  are extracted as candidate words for  $a_2$ .

For one of the target pair ( $i_1$ ) (specifically, the word related to the focusing feature selected by the user), the top 10 words with the highest relevance scores ( $r_{i_1j}$ ) are extracted as the base concept ( $a_1$ ). Subsequently, for each base concept, the top 10 words with the highest cosine similarity to  $\hat{v}(a_2)$  are retrieved as the corresponding concept ( $a_2$ ) of the base concept to create the base pair ( $a_1, a_2$ ). This approach generated 100 base pairs in total. Feature extraction is then performed on the base pairs using the method described for feature extraction. Among these base pairs, those sharing more features with the target pair are considered more appropriate. For multiple pairs sharing the same number of common features, the pair with higher cosine similarity is considered more suitable. Finally, the top 30 base pairs in terms of suitability were presented to the user for selection.

### 3.5 Explanation Generation

The LLM generates an explanation using a prompt that combines the given target pair, focusing feature, and base concept or pair. For example, when the target pair is (“electric current,” “voltage”), the feature is “electric current flows,” the base concept is “water,” and the base pair is (“water,” “water pressure”), the prompt using the base concept or the base pair in Japanese is formulated as follows.

Prompt using the base concept: Focusing on “electric current flows” for “electric current” and “voltage”, provide an explanation by using the analogy that compares “electric current” to “water”.

Prompt using the base pair: Focusing on “electric current flows” for “electric current” and “voltage”, provide an explanation by using the analogy that compares “electric current” to “water” and “voltage” to “water pressure”.

The LLM then generates an explanation based on the provided analogy.

## 4 Evaluation Experiment

### 4.1 Experiment1

An evaluation experiment was conducted to determine the optimal value of the parameter  $\alpha_c$  in

Equation (1), which represents the weight assigned to the user-selected focusing feature.

#### 4.1.1 Method

Table 1 shows the queries of the target pairs and features used in the experiment. For each of the parameter values  $\alpha_c = 1.0, 2.5, 5.0, 7.5, 10.0$ , the top 10 base concepts with the highest relevance scores  $r_{ij}$  were generated. The base concepts generated for each target pair were divided into eight subsets to create eight distinct question sets. Participants were asked to evaluate the appropriateness of each base concept as a metaphorical expression for the target pair and focusing feature, using the following response options: “highly appropriate,” “appropriate,” “neutral,” “inappropriate,” “completely inappropriate,” and “do not understand the meaning of the word.” Each set contained approximately 100 items. The experiment was conducted using a crowdsourcing platform. For each set, responses were collected from 100 to 105 participants, totaling 816 participants across all eight sets.

#### 4.1.2 Result

For each of the six response options, numerical values were assigned as follows: 2 for “highly appropriate,” 1 for “appropriate,” 0 for “neutral,” -1 for “inappropriate,” -2 for “completely inappropriate,” and -2 for “do not understand the meaning of the word.” For each analogical concept, the mean of these scores—weighted by the reciprocal of the appropriateness rank based on the relevance score  $r_{ij}$  of each parameter—was calculated as the evaluation score. For each target pair and feature (query), the parameter with the highest evaluation score was regarded as the parameter that produced the optimal estimation for each query. The number of queries for which each parameter was optimal is shown in Table 2.

#### 4.1.3 Discussion

The highest evaluation score was obtained when the parameter value was 2.5. However, the parameter value that yielded the highest evaluation score for the largest number of sets was 1.0, followed by 10.0. These results suggest that the optimal parameter values may vary depending on the set. Thus, rather than fixing the parameter to a single value, the system allows users to adjust the parameters while selecting base pairs.

Table 1: Target pairs and focusing features

| Target pair               | Feature                                                                                                                                                     |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Guitar, Amplifier         | Combine amplifier<br>Connect amplifier                                                                                                                      |
| Nucleus, Electron         | Electron revolves<br>Electron exists                                                                                                                        |
| Trial, Juror              | Conduct trial<br>Deliver at trial<br>Capable of conducting trial                                                                                            |
| Antibiotic, Bacteria      | Kill bacteria<br>Infected with bacteria<br>Detect bacteria                                                                                                  |
| Journal, Paper            | Read paper<br>Submit paper<br>Publish paper<br>Published in journal                                                                                         |
| Electric Current, Voltage | Generate electric current<br>Measure electric current<br>Electric current flows<br>Proportional to voltage<br>Measure voltage                               |
| Levee, Flood              | Build levee<br>Overflow levee<br>Prevent flood<br>Cause flood<br>Flood occurs                                                                               |
| Cabinet, Prime minister   | Elect prime minister<br>Prime minister performs<br>Belong to cabinet<br>Incumbent upon cabinet<br>Comprise cabinet                                          |
| Function, Argument        | Return function<br>Define function<br>Declare function<br>Pass to function<br>Function executes<br>Pass argument<br>Receive an argument<br>Take an argument |

## 4.2 Experiment2

The analogical explanations generated by the system were experimentally validated.

### 4.2.1 Method

Based on the results of Experiment 2, base pairs were selected from the candidates generated using parameter values of 1 or 10 for each query comprising the target pair and focusing feature. The target pairs ( $i_1, i_2$ ), focusing features ( $feature$ ), and base pairs ( $a_1, a_2$ ) consisting of base concepts ( $a_1$ ) and corresponding concepts ( $a_2$ ) used in the experiment are summarized in Table3.

Using OpenAI ChatGPT 3.5, explanations of

Table 2: Average evaluation scores (Ave. ES) and the number of queries for which the parameter was optimal (Num. of OPV)

|             | parameter $p_c$ |             |      |      |      |
|-------------|-----------------|-------------|------|------|------|
|             | 1.0             | 2.5         | 5.0  | 7.5  | 10.0 |
| Ave. ES     | 0.19            | <b>0.20</b> | 0.18 | 0.14 | 0.02 |
| Num. of OPV | <b>18</b>       | 2           | 5    | 4    | 8    |

approximately 200 Japanese characters were generated to eliminate the influence of explanation length, employing prompts in Japanese corresponding to the following four conditions:

- **Target pair only condition**  
For  $i_1$  and  $i_2$ , please provide an explanation of approximately 200 characters by using an analogy. (Because there is no qualitative distinction between  $i_1$  and  $i_2$ , their order was determined by the order of input.)
- **Feature condition**  
Focusing on *feature* for  $i_1$  and  $i_2$ , please provide an explanation of approximately 200 characters by using an analogy.
- **Base concept condition**  
Focusing on *feature* for  $i_1$  and  $i_2$ , please provide an explanation of approximately 200 characters by using the analogy that compares  $i_1$  to  $a_1$ .
- **Base pair condition**  
Focusing on *feature* for  $i_1$  and  $i_2$ , please provide an explanation of approximately 200 characters by using the analogy that compares  $i_1$  to  $a_1$  and of  $i_2$  to  $a_2$ .

The analogies were generated under the target pair only and feature conditions are shown in Table4. Depending on the type of query, the generated explanation involved either the base pair ( $a_1, a_2$ ), only the base concept ( $a_1$ ), the corresponding base concept ( $a_2$ ) that analogizes the target concept unrelated to the focusing feature, or the base domain encompassing the base pair.

Participants were asked to rate the generated explanations on a five-point scale (assigned numerical values from 1 to 5) in terms of explanation clarity, appropriateness of the analogy, appropriateness of the feature for the target pair, and appropriateness of the explanation focusing on the feature. Under the target pair only condition, the two feature-related appropriateness ratings were not obtained.

Table 3: Target pairs, focusing features, base concepts and the pairs used in Experiment2

| Target pair            | Focusing feature            | Base pair<br>(Base concept, corresponding concept) |
|------------------------|-----------------------------|----------------------------------------------------|
| Flood, Levee           | Flood prevention            | Wave, Seawall                                      |
| Levee, Flood           | Overflowing levee           | Road, Traffic jam                                  |
| Paper, Journal         | Publish paper               | Manuscript, Newspaper                              |
| Journal, Paper         | Publish in journal          | TV, Program                                        |
| Prescription, Pharmacy | Accept prescription         | Document, Government office                        |
| Pharmacy, Prescription | Go to pharmacy              | Shop, Order                                        |
| Amplifier, Guitar      | Connect amplifier           | PC, Keyboard                                       |
| Equation, Math         | Handle equation             | Law, Japanese language                             |
| Government, Law        | Give to government          | Organization, Rule                                 |
| Cell, Biology          | Composed of cell            | Organization, Society                              |
| Earthquake, Epicenter  | Earthquake occur frequently | Incident, Origin                                   |
| Sample, Statistics     | Extract sample              | Model, Plastic model                               |

Table 4: Analogies supplied by LLM. Pair: base pair ( $a_1, a_2$ ). Unmarked: base concept  $a_1$ . \*: corresponding base concept  $a_2$ . \*\*: base domain

| Target pair            | Analogy                        |                               |
|------------------------|--------------------------------|-------------------------------|
|                        | Target pair only               | Feature                       |
| Flood, Levee           | Giant barrier (control valve)* | Castle walls*                 |
| Levee, Flood           | Giant barrier (control valve)  | Shield                        |
| Paper, Journal         | Recipes, Cookbook              | Art pieces, Gallery           |
| Journal, Paper         | Cookbook, Recipes              | Cookbook, Recipes             |
| Prescription, Pharmacy | Pathfinder, Adventure guide    | Magic spells                  |
| Pharmacy, Prescription | Adventure guide, Pathfinder    | Puzzle keys*                  |
| Amplifier, Guitar      | Ink, Brush pen                 | Cooking utensils, Ingredients |
| Equation, Math         | Compass                        | Adventurer’s tools            |
| Government, Law        | Captain, Voyage design         | Captain                       |
| Cell, Biology          | Bricks and Beams, Building     | Buildings and road networks   |
| Earthquake, Epicenter  | Pressure cooker**              | Stage**                       |
| Sample, Statistics     | Waves, Sea                     | Art museum**                  |

Four sets of questionnaires were constructed such that items concerning the four types of explanations generated under different conditions for the same query did not overlap within each set. Each set included items related to explanations generated for three distinct queries under each condition. An experiment was conducted using a crowdsourcing platform. For each set, responses were collected from 107 to 142 participants, totaling 504 participants across all eight sets.

#### 4.2.2 Result

The average ratings for the explanation clarity are shown in Figure2-a. A one-way ANOVA with corrections for violations of sphericity was conducted, and the results reveal a significant main effect of the prompt condition ( $F(2.96, 1490.35) = 36.18, p < .01, \eta^2 = .0360$ ). Multiple comparisons using Holm’s method show significant differences at the 1% level for all pairs except between the target pair only condition and feature condition, and between the target pair only condition and base pair condition.

The average ratings for the appropriate-

ness of the analogy are shown in Figure2-b. The ANOVA results show a significant main effect ( $F(3, 1509) = 26.50, p < .01, \eta^2 = .0261$ ). Multiple comparisons show significant differences at the 1% level for all pairs except between the target pair only and base pair conditions.

The average ratings for the appropriateness of the feature for the target pair are shown in Figure2-c. The ANOVA results indicate a significant main effect ( $F(1.97, 990.97) = 34.16, p < .01, \eta^2 = .0272$ ). Multiple comparisons show significant differences at the 1% level for all pairs.

The average ratings for the appropriateness of the explanation focusing on the feature are shown in Figure2-d. The ANOVA results show a significant main effect ( $F(2, 1006) = 45.89, p < .01, \eta^2 = .0400$ ). Multiple comparisons reveal significant differences at the 1% level for all pairs.

#### 4.2.3 Discussion

The results of the experiment show that the base concept condition received the highest ratings across all four evaluation criteria. The lower ratings observed in the target pair only and feature condi-

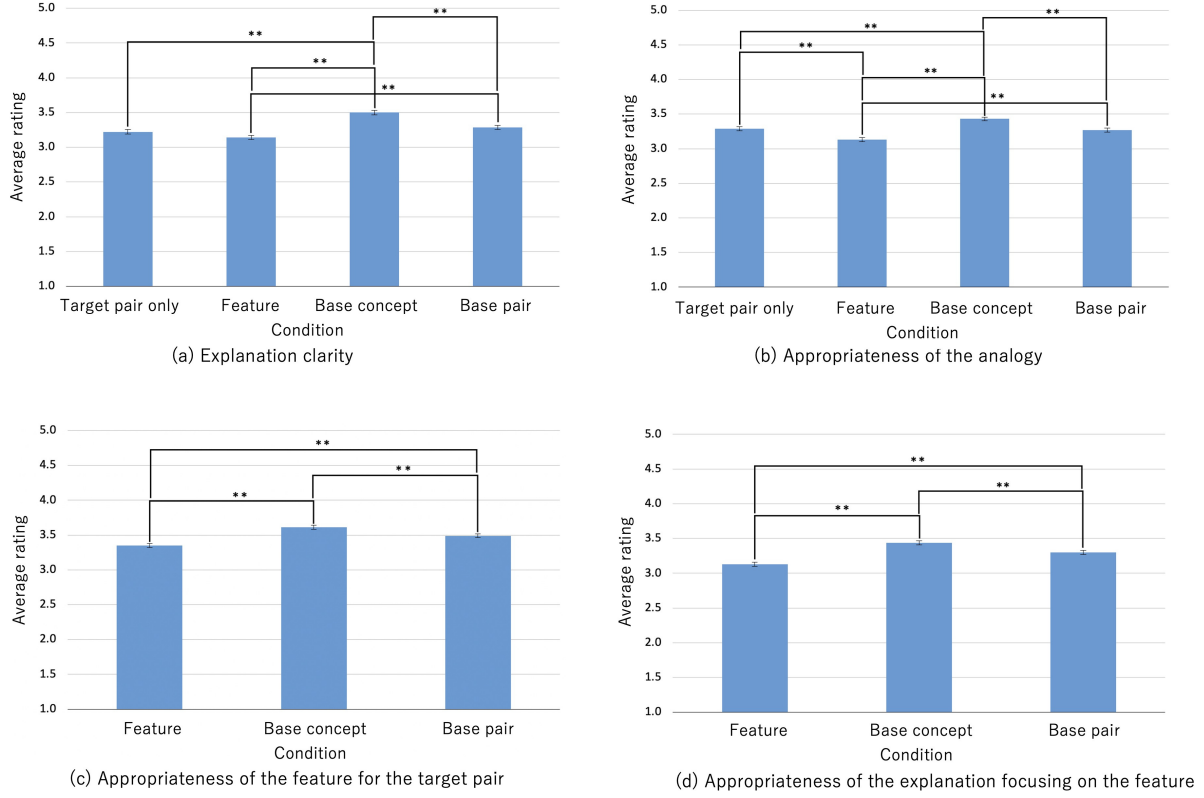


Figure 2: Results of Experiment 2. (a) Average ratings of the explanation clarity (b) Average ratings of appropriateness of the analogy (c) Average ratings of appropriateness of the feature for the target pair (d) Average ratings of appropriateness of the explanation focusing on the feature

tions, where no concepts for analogy were provided, suggest that the system-supplied base concepts for analogy were more suitable for generating analogical explanations than the LLM-supplied concept.

However, no substantial differences were observed between the target pair only and base pair conditions. This finding suggests that the four-term analogy approach may not perform satisfactorily. The four-term analogies were generated using an existing method based on embedding vectors without modification, and the evaluation method for the generated base pairs may have been insufficient.

Furthermore, significant differences were observed between the conditions for the appropriateness of features with respect to the target pairs. Similar to the other evaluation criteria, the base concept condition received the highest ratings. This is likely owing to the appropriateness of the explanation based on the features. Even when the features are identical, their evaluation is affected by how accurately they are explained.

## 5 Conclusion

In this study, we proposed a human-in-the-loop system that generates explanations for two related concepts by acquiring appropriate base concept pairs for analogy and using an LLM to produce explanations. The results of the evaluation experiment indicated that explanations generated using prompts containing only base concepts achieved the highest ratings, demonstrating the validity of extracting base concepts using the proposed system. In contrast, prompts providing base pairs did not perform as well, indicating that it is necessary to improve the four-term analogy generation process of the system. In this study, GPT-3.5 was used, and advances in LLM technology may enable improved results through the use of more advanced LLM (e.g. GPT5.0). However, the results of this study suggested that a human-in-the-loop framework leads to more appropriate outputs in analogy generation.

In addition, experiments on the parameters used to extract base concepts based on user-selected focusing features showed that the optimal parameter values depend on the target pair and feature, suggesting that users must adjust these parameters in

the proposed system. Identifying the appropriate parameter conditions for each query comprising the target pair and feature could potentially enable the automatic adjustment of parameter values. However, the experiments mainly assessed the appropriateness of system-generated base concepts. Therefore, additional experiments are necessary to evaluate the entire interactive process from beginning to end.

Finally, regarding the prompt design, this study provided minimal instructions to the LLM to evaluate the effectiveness of the system-supplied analogies. The results confirmed the effectiveness of prompts that include analogical concepts provided by the system. Future studies should investigate the effects of varying the format of prompts containing analogies on the generated explanations to identify the optimal prompt structure.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research(C) 24K15073. This manuscript was edited by editage (Editing Service).

## References

- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.
- Zied Bouraoui, José Camacho-Collados, and Steven Schockaert. 2019. Inducing relational knowledge from BERT. In *AAAI Conference on Artificial Intelligence*, pages 7456–7463.
- Ryoichi Chikamura, Rafal Rzepka, and Kenji Araki. 2016. A method of collecting figurative explanations from the web[in Japanese]. In *Proc. of 27th Fuzzy System Symposium*, pages 112–115.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: pre-training of deep bidirectional transformers for language understanding. In *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170.
- Dedre Gentner and Donald R. Gentner. 1983. Flowing waters or teeming crowds: Mental models of electricity. *D. Gentner, A. L. Stevens (Eds.), Mental models*, pages 99–129.
- Syuichi Ishida, Takahiro Kurakata, and Tatsunori Mori. 1996. Metaphor generation for explanation in intelligent manual systems[in Japanese]. *Special Interest Group on Natural Language Processing (SIG-NL)*, pages 1–6.
- Daisuke Kawahara and Sadao Kurohashi. 2016. Case frame compilation from the web using high-performance computing. In *5th edition of Intl. Conf. on Language Resources and Evaluation*, pages 1–13.
- James C. McCroskey and Walter H. Combs. 1969. The effects of the use of analogy on attitude change and source credibility. *Journal of communication*, 19(4):333–339.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint, arXiv:1301.3781*.
- Ilyse Resnick, Alexandra Davatzes, Nora S. Newcombe, and Thomas F. Shipley. 2016. Using analogy to learn about phenomena at scales outside human perception. *Cognitive Research: Principles and Implications*, 5:1344–1347.
- Lindsey E. Richland, Osnat Zur, and Keith J. Holyoak. 2007. Cognitive supports for analogies in the mathematics classroom. *Science*, 316(5828):1128–1129.