Proposal of a Semi-Automatic Classification Method for Estimating Conceptual Understanding in Short Answer Grading for Semi-Open-Ended Questions Using Word Co-occurrence Networks

Katsuko T. Nakahira
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

***Abstract*— One of the effective method for estimating learners' level of understanding of acquired concepts involves using semi-open-ended questions. The method is well known to be particularly beneficial for questions requiring scientific explanations, and it is adopted in large-scale academic achievement tests and trend assessments. On the other hand, Short Answer Grading often relies on manual scoring by markers, raising concerns about workload and the depth of insight into specialized knowledge. To alleviate marker workload, automated evaluation combining natural language processing and machine learning, or utilizing generative AI, is anticipated. However, introducing machine learning requires large amounts of training data and also faces issues related to the native language of the test takers. As one solution to these problems, a method that enables classifying learners' understanding levels with minimal effort based on collected short answers, even under conditions of limited information for machine learning, is also considered beneficial. In this paper, we propose a method that classifies short answers into three levels based on conceptual understanding depth as one approach to short answers grading for semi-open-ended questions. The proposed method applies degree analysis, well-known in word co-occurrence networks.**

Keywords-Semi-Open-Ended question; short answer grading; word co-occurrence network; degree analysis.

I. INTRODUCTION

With the advancement of natural language processing and generative Artificial Intelligence (generative AI), many researches in recent years have focused on the automated evaluation of descriptive responses to open-ended question.

For questions requiring scientific explanations, it is preferable to use a Short Answer format for semi-open-ended problems, and the marker should grade them manually. In this paper, the term "semi-open-ended" answer refers to response that, while open-ended, are subject to certain constraints on the perspective taken—such as an interpretation of trends in a statistical graph. While the usefulness of large-scale academic achievement tests and trend tests has been noted, the burden on markers due to the enormous number of examinees and the depth of insight required for specialized knowledge have long been pointed out. This marker-based grading, known as Short Answer Grading (SAG), involves assessment to measure student learning outcomes. However, in today's world where student numbers or their growth increases exponentially, this task has become a difficult challenge for educators. To address the above issues, Automated SAG (ASAG) systems combining natural language processing and machine learning, as well as ASAG systems utilizing generative AI, have been implemented.

Analysis of text sentiment in natural language processing has long been integrated with knowledge from network science, yielding various practical studies in the form of Word Co-occurrence Networks (WCNs). In recent years, research into the structure and dynamics of complex networks has attracted the attention of both academic researchers and practitioners.

Applications of the WCN method to long text include the following examples. Some researcher created WCNs for academic papers published in the field of information bibliometrics (see Sedighi [1]) or bioinformatics (see Li et.al. [2]) and analyzed their structural features and depth. These studies confirmed small-world properties and provided insights into temporal transitions of concepts and word usage. Additionally, Liang et al. [3] constructed WCNs for standard English articles, calculated the spectra of their adjacency matrices, and identified characteristic distributions in the spectral density distribution through qualitative analysis of the spectra.

Research applying WCN to short text corpora includes, for example: Garg et al.'s [4] application to microblogs, Fudolig et al.'s [5] application to political tweets, and Amancio et al.'s [6] application to short texts. In each case, the structure of WCNs differed from that of longer texts, suggesting the existence of characteristic parameters and the necessity of specific analytical methods.

Additionally, there is a research example focusing on the context of text authors, Millington and Luz [7]. Using the ADReSS Challenge dataset, a natural speech dataset, they constructed WCNs from conversation records of healthy individuals and potential Alzheimer's disease patients and compared structural metrics. The results revealed distinct features in heterogeneity, centrality, and edge density specific to the networks of Alzheimer's patients, with many network characteristics being driven by word frequency. This suggests that, overall, classification between the healthy control group and the Alzheimer's group is possible based solely on the structure of the WCN.

Proposals for ASAG systems combining natural language processing and machine learning include, for example, the following: A method for separating essay groups into subsets representing similar graders using explanatory variables and clustering (Zupanc and Bosnić [8]), a method utilizing rubric information (Wang et al. [9]), a method that introduces a Transformer encoder layer into a BERT model and trains its weights solely on the text of the rubric criteria table (Condor et al. [10]), focusing on semi-open-ended short-answer questions and integrating both general domain knowledge and domain-specific information (Zhang et al. [11]), and one that utilizes very basic natural language processing techniques such as N-grams and word embedding technology along with machine learning algorithms (Lertchaturaporn and Pokpong [12]), among others. Additionally, there are investigations into the effects on learners' trust in grading and learning motivation (Conijin et al. [13]). At present, ASAG systems centered on natural language processing and machine learning are mainstream, though there are many applications of ASAG systems using generative AI (Jamil and Hameed [14], Chang and Ginter [15], Grévisse [16], Mello et al. [17]).

Taking these trends in consideration, future possibilities for ASGA include not only focusing on accurate grading but also categorizing response levels and conducting detailed reviews as needed for assigned levels. Particularly when using machine learning, the one of greatest effort lies in the necessity of training the system using past answer patterns and their corresponding grading results. Additionally, working language problem adopting toward machine learning issues arise, especially for educational institutions that accept many international students. Of course, to reduce the effort required for grading, the use of machine learning or generative AI is ultimately desirable. However, as an alternative approach, even if it doesn't reduce effort as much as machine learning, a mechanism that allows classifying learners' level of understanding with minimal effort based on collected short answers—even in situations where information for machine learning is scarce—is worth considering. Based on this idea, we propose a framework for educational settings to classify learners' short answers to semi-open-ended questions—even with minimal knowledge—into three levels: 1) Simple recall, 2) Explanation through reconstruction of learned knowledge, 3) Application of learned knowledge. Through this framework, there is a possibility to categorize learners' comprehension levels and knowledge expression during lectures or briefings, thereby supporting subsequent educational efforts.

This paper is structured as follows. Section 2 describes the estimation of WCN structures based on the SAG process for semi-open questions, using the concept of degree in complex networks. Section 3 introduces the deep structure of the nearest-neighbor plane based on the framework described in Section 2. It then discusses key parameters for estimating this deep structure by performing a simplified simulation on it. We conclude the paper in Section IV.

II. Short Answer Generating Process for Semi-Open-Ended Question and Word Co-Occurrence Network

A. *Selecting the Words for Generating Answer for Semi-Open-Ended Question*

Examples of using semi-open-ended questions include taking minutes after a lecture or reading comprehension exercises when figures or tables are presented. A common purpose for adopting semi-open-ended questions is to assess learners' level of understanding when they possess specific prior knowledge. Here, we describe the cognitive model involved in solving semi-open-ended questions, specifically the process of generating semi-open-ended answers.

First, learners acquire newly presented concepts as knowledge. This knowledge is, in principle, encoded in some form. In the context of school learning, concepts are fundamentally encoded as language. These encoded concepts are then stored in long-term memory as either technical terms used in specific contexts, general terms used to explain those technical terms, or linked to other technical terms. In this process, the language used to explain the concept is stored in long-term memory by linking several words together.

This shares similarities with the micro structure of Generic Skills introduced by Nakahira et al. [18]. That paper introduced a framework where Generic Skills consist of three stages: perception of physical contents provided through lectures or experiments, collection of perceived contents, and reconstruction. Collected contents become information, and when information is organized, it becomes knowledge. The series of activities suggested that learner behavior differs depending on whether the acquisition action is passive or active. Specifically, it was introduced that in passive acquisition, transitioning collected content beyond simple memorization is difficult, whereas in active acquisition, further exploration of the obtained information and application to similar situations enable transition to the re-representation of acquired knowledge.

In this paper, we apply a series of micro structures to the act of solving semi-open-ended questions. Short answers observed when solving semi-open-ended questions regards as reflecting one of the following three states, depending on the level of mastery of the learned concepts.

- **Level A: Simple Memorization of Learned Knowledge**
  Reciting learned concepts verbatim (simple memorization)
- **Level B: Explanation through reconstruction of learned knowledge**
  Able to explain learned concepts by substituting similar words for the original terms (paraphrasing, somewhat deeper understanding)
- **Level C: Application of learned knowledge**
  Able to use learned concepts to explain similar concepts or phenomena (application of knowledge)

Hence, the vocabulary used in the short answers written in the responses can be also considered to reflect the above proficiency levels.

The phrases selected for short answer writing can be observed as multiple words appearing simultaneously, i.e., in the form of word co-occurrence. In the field of natural language processing, this has traditionally been addressed using Word Co-occurrence Networks (WCNs). This paper adapts this idea, proposing that the selection of word groups used in short answer descriptions and learners' proficiency levels regarding

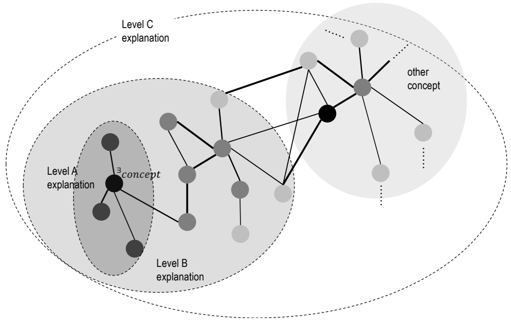

Figure 1. Conceptual diagram of the network of words used to explain a certain concept.

learning concepts are reflected in WCN features. We also propose analyzing this using the following framework.

B. Description of the Relationship between Degree and Conceptual Proficiency Levels for WCN Words Appearing in Short-Answer Questions

For a given word W_i that we adopt when generating text, we set the following assumption. The contexts in which W_i is used can be divided into two categories:

- General words, which are used without reference to a specific context.
- Specific words, which are only used in specific contexts.

Now, we imagine that someone are creating a short answer for a given semi-open-ended question. Since the question is a semi-open-ended given in a proficiency assessment setting, the short answer will contain a mix of two types of descriptions: the precise knowledge expected and descriptions that further develop that knowledge through reasoning. In this process, we assume that descriptions of the concept will be formed through the following steps, depending on the level of understanding (proficiency level) of the knowledge acquired in the past.

- **Level A**
  The concepts learned are explained using specialized words, so terms not commonly used in everyday life are frequently employed. In other words, specific words appears frequently.
- **Level B**
  After simple memorization, the learned knowledge is explained by reconstructing the concepts. Therefore, even when using technical terms, they are simultaneously replaced with alternative expressions. The words adopted at this time are not necessarily technical terms. That is, general words appear mixed among specific words.
- **Level C**
  When explaining similar concepts or phenomena, the application is not necessarily specialized, so technical terms are used alongside general words in short answers. That is, specific words appear among many general words.

The selection of words used to explain a concept depends on the words associated with that concept. This conceptual diagram is shown in Figure 1. When first learning a concept, one can only explain it using the exact term taught. Therefore, explanations will likely be limited to words found within the very narrow scope of Level A. As understanding of the concept broadens slightly, it becomes possible to replace the explanation with words from Level B, using words that can connect to areas outside Level A. With further deepening of understanding, it becomes possible to use words connecting to other concepts to provide explanations that link the concepts together. This state is considered to be Level C.

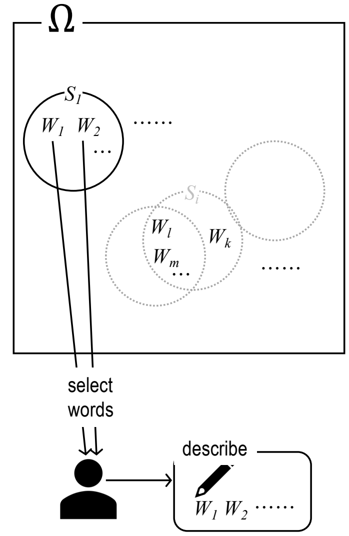

Figure 2. The relation between words selection and short answer writing.

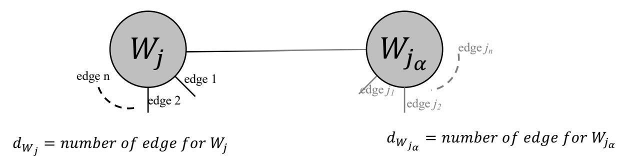

Figure 3. Relationship between nodes and edges for W_j and W_{j_α}.

When considering the “word network” during generating short answer, the process follows the flow described above, and this process can related to human cognitive behavior. Figure 2 shows a conceptual diagram of the word selection process when learners write short answers after being presented with a semi-open-ended question. In the case of semi-open-ended questions administered after a lecture, they are typically used by instructors to assess learners’ proficiency levels. Therefore, it is assumed that semi-open-ended questions are set to multiple learners, typically tens to hundreds of individuals. After the lecture concludes, or in some other context, when an instructor provides a semi-open-ended question to learners, learner i creates and submits a short answer S_i. As a result, after the semi-open-ended question is completed, short answers $\boldsymbol{S}$ from n_p respondents can be collected.

$$\boldsymbol{S} = \{S_i(\boldsymbol{W}_i(W_j)) \mid i = 1, \cdots, n_p,\ j = 1, \cdots, n_{w_i}\}.$$

S_i contains n_{w_i} words, and learners express concepts learned by linking multiple W_j.

The number of words W_j can link, i.e., the number of edges when W_j is a node, is called the degree which is denoted by d_{W_j}. The relationship between W_j, d_{W_j}, and the associated

TABLE I. TREND OF DEGREE PAIRS $(d_p,\ d_q)$ FOR WORD PAIRS $W_p,\ W_q$. "C" DENOTES COMPLEX, "A" DENOTES AVERAGE, "S" DENOTES SPARSE.

		degree for word p		
		large	moderate	small
degree for word q	large	c/c	a/c	s/c
	moderate	c/a	a/a	s/a
	small	c/s	a/s	s/s

W_{j_α} and $d_{W_{j_\alpha}}$ is shown in Figure 3. In this context, d_{W_j} can be interpreted as follows: if W_j is a global word, it can link to many words, meaning large d_{W_j}; if it is a specific word, it can link only to a few words, meaning small d_{W_j}. Therefore, this situation suggests that by performing degree analysis on W_j, it may be possible to investigate a learner's proficiency level regarding specific concepts they have learned, even before confirming the content of the text. This is described as follows.

Let $\mathbf{\Omega}$ denote the set of all unique words W_α appearing in $\boldsymbol{S}$.

$$\mathbf{\Omega} = \{W_\alpha \mid \alpha = 1, \cdots, n_{w_t}\}$$

Here, n_{w_t} is the number of the set. Each W_α take a value of d_{W_α}.

When the learners generate short answers for semi-open-ended questions, they extract n_{w_i} words from $\mathbf{\Omega}$ and concatenate them in an appropriate order. In this case, the word(s) linked to W_j each possess a degree. To grasp the general trend before reading the set of short answers, examining the degree of the word(s) linked to W_j allows for a simple check, as shown in Table I. Table I shows that when considering a word pair $(W_p,\ W_q)$, the combination of the degrees of the degree pair $(d_p,\ d_q)$ results in expressions ranging from c/c (simple explanations using words not tied to specific contexts) to s/s (complex explanations using words tied to specific contexts).

Since W_j can connect to d_{W_j} words, there are d_{W_j} possible combinations of $(W_p,\ W_q)$. Given this characteristic, it would be beneficial to have a method that simultaneously represents the properties of W_j and the properties of all words linked to W_j. As an idea, treating the number of words W_j itself can link to and the total number of words connected to W_j that can link to other words as features would enable efficient analysis. We describe the method as follows.

d_{W_j} represents the number of words W_j can link to other words, or in other words, the number of first-order nearest neighbors. This is denoted as $d_{j_{NN_1}}$. $d_{j_{NN_1}}$ suggests it can serve as a criterion for determining whether W_j is a general word or a specific word.

Next, consider the characteristics of the words $W_{j_\alpha}(\alpha = 1, 2, \cdots, d_{W_j})$ connected to W_j. Like W_j, W_{j_α} also has degree d_{j_α}. The property of words that can link to W_j, namely their degree of general/specific words, is expressed as the total number of nodes linked to W_j that possess d_{j_α}. This corresponds to determining the number of second-order nearest neighbors from the perspective of W_j. We denote this as $d_{j_{NN_2}}$, and described by the following equation.

$$d_{j_{NN_2}} = \sum_{\alpha=1}^{d_{j_{NN_1}}} (d_\alpha - 1) \tag{1}$$

The reason for subtracting 1 from d_α is to prevent double counting of connections, since W_{j_α} is originally linked to W_j.

The $(d_{NN_1},\ d_{NN_2})$ calculated as described above are represented numerically, allowing their occurrence frequency to be examined. If the occurrence frequency of a specific $(d_{NN_1},\ d_{NN_2})$ is high, we can consider that $\boldsymbol{S}$ is expressed as a phrase possessing a specific combination of word properties. Therefore, it suffices to find a method to reproduce the distribution of $(d_{NN_1},\ d_{NN_2})$.

III. Reproduction of Degree Distribution through Superposition of Probability Density Distributions

The frequency distribution of $(d_{NN_1},\ d_{NN_2})$ shown in the previous chapter is the distribution of WCN pairs NN_1 and NN_2 derived from human-generated text. There are many various classification methods and metrics within complex networks to investigate this tendency. This paper posits that the frequency distribution of $(d_{NN_1},\ d_{NN_2})$ follows some probability distribution, as it is generated in accordance with human thought. While numerous types of probability distributions exist, this paper assumes a two-dimensional normal distribution for simplicity.

The two-dimensional normal distribution for variables x and y is described by the following equation.

$$\begin{aligned} p(x,y) &= \frac{1}{R}\exp\left(-\frac{1}{2(1-\rho^2)}(X^2+Y^2-C)\right), \\ \text{here,} & \\ R &= 2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}, \\ X &= \frac{x-\mu_x}{\sigma_x}, \\ Y &= \frac{y-\mu_y}{\sigma_y}, \\ C &= \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}. \end{aligned}$$

When mixing $p(x,\ y)$, let K denote the number of mixture components. The actual distribution $F(d_{NN_1},\ d_{NN_2})$ of $(d_{NN_1},\ d_{NN_2})$ can be described as follows.

$$F(d_{NN_1},\ d_{NN_2}) = \sum_k^K \pi_k p_k(x,\ y),\ \text{here,}\ \sum_k \pi_k = 1$$

For $F(d_{NN_1},\ d_{NN_2})$, considering a specific mixture component k, the following exist for $p_k(x,\ y)$: mean $\mu_x,\ \mu_y$, standard deviation $\sigma_x,\ \sigma_y$, correlation coefficient ρ. These are determined for each NN_1 and NN_2 belonging to k. Furthermore, for each p_k, the mixing weight π_k should ideally also be considered.

Using the above parameters and equations, in this paper we set $K = 2$ to create a template matching the distribution of $F(d_{NN_1},\ d_{NN_2})$. Then we discuss the tendency of $\boldsymbol{S}$ when a distribution resembling this template is observed. Table II shows the parameters we used, and calculations were performed using various combinations. Since p_k includes

TABLE II. THE PARAMETER SET USED FOR COMPUTING THE PROBABILITY DENSITY DISTRIBUTION. THE COMPUTATION RANGE IS FROM -5 TO 5.

parameter	value set
μ_{NN_1}	$-3,\ -1,\ 0.5,\ 1,\ 3$
μ_{NN_2}	$-3,\ -1,\ 0.5,\ 1,\ 3$
σ_{NN_1}	$1,\ 2,\ 3$
σ_{NN_2}	$1,\ 2,\ 3$
ρ	$0.1,\ 0.25,\ 0.5,\ 0.75,\ 0.9$
π_k	0.5

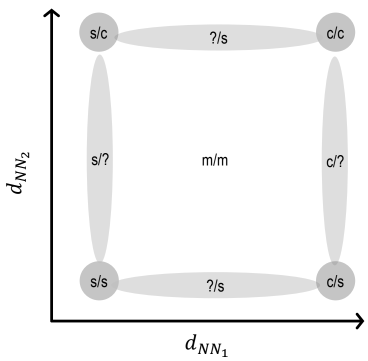

Figure 4. Prediction of the occurrence of each state in Table 1 using mixed synthesis. The gray shaded area indicates the approximate distribution.

$\mu_{NN_1},\ \mu_{NN_2},\ \sigma_{NN_1},\ \sigma_{NN_2},\ \rho$, the parameters for each k were calculated by combining the values from Table II and summing them with each weight π_k set to 0.5. Figure 5, Figure 6, and Figure 7 show the computational results. Based on the results, we discuss about the behavior. All figures display the probability density at its maximum value. When calculating the probability density distribution, the computational range for both NN_1 and NN_2 is from -5 to 5.

A. Distribution of $(d_{NN_1},\ d_{NN_2})$ under the Mixture Model

The template created as described above, which mimics the $(d_{NN_1},\ d_{NN_2})$ distribution, is represented as a density distribution on a plane with d_{NN_1} on the horizontal axis and d_{NN_2} on the vertical axis, as shown in Figure 4. In this paper, we refer to this as the *depth structure of nearest neighbor*. Interpreting this together with Table I, the regions to consider are those shown in the figure: c/c, c/s, s/c, s/s, and regions with one side fixed.

Here, we interpret the meaning of vertical axis direction and change value. From (1), d_{NN_2} is the sum of the degrees of all words connected to W_j. This means that equation (1), by taking the sum of d_{j_α} for connected words, emphasizes the properties of the group of words that W_j can combine with.

When d_{NN_2} has a large value, W_j indicates that it has many links with global words, meaning it is a more general-purpose word. Concepts explained by using many general-purpose words generally correspond to explanations using examples or analogies, or descriptions based on deeper understanding by linking to other concepts. In Figure 4, a strong peak in the c/c region indicates the presence of careful explanations using examples. A strong peak in the s/c region signifies that the specific word itself represents a concept capable of having many connections, and that it is being explained more concretely using technical terms and examples. In either case, it indicates the presence of descriptions that utilize newly acquired knowledge.

When d_{NN_2} has a small value, W_j has only links to specific words. This means that even if W_j were a global word, it would only be used to explain specific words. That is, it indicates that after learning the concept of a new item, learners attempted to explain it using the exact learned phrase without much elaboration. This tendency is commonly observed immediately after learning the concept of a new item. Particularly when a strong peak appears in the s/s area in Figure 4, it is reasonable to assume that many learners absorbed the new item but remain in a state of mere absorption.

Considering the above points comprehensively, the distribution of $(d_{NN_1},\ d_{NN_2})$ enables estimation of temporal transitions when administering similar tasks after a period following training. Although a strong peak is observed in the s/s region immediately after learning, as learning progresses, it changes in the vertical or horizontal direction. If it transitions to c/c via s/c or c/s, it indicate that standard learning deepening has been achieved.

In the case of analysis toward actual short answer data, we can easy image that such distributions do not necessarily appear in a single specific location. Therefore, we consider synthesizing K probability density functions and estimating learners' acquisition of new concepts by determining where the composite result shows strong responses.

B. The Effect of the Standard Deviation in p_k on the Mixture Model

Figure 5 shows the computation results for the influence of $\sigma_{NN_1},\ \sigma_{NN_2}$ for each p_k on the mixture model distribution. Here, $\mu_{NN_1},\ \mu_{NN_2}$ for p_1 were set to $(3,\ 3)$, $\mu_{NN_1},\ \mu_{NN_2}$ for p_2 as $(3, 1)$, and ρ fixed at 0.5. $\sigma_{NN_1},\ \sigma_{NN_2}$ were varied from 1 to 3. In the figure, the top row shows the mixture model distribution when $\sigma_{NN_1} = \sigma_{NN_2}$, the middle row shows the distribution when σ_{NN_1} for p_1 is set to 1 and the other standard deviations are varied, the bottom row shows the distribution when σ_{NN_1} for p_1 is set to 2 and the other standard deviations are varied.

The upper right section shows that both p_1 and p_2 have $(\sigma_{NN_1},\ \sigma_{NN_2}) = (3,\ 3)$. The distribution is a diffuse witout specific concentration large value. Therefore, we concluded that creating a template is unnecessary when the standard deviation is excessively large, and thus created combinations with standard deviations of either 1 or 2.

Overall, when p_k includes $(\sigma_{NN_1},\ \sigma_{NN_2}) = (1,\ 1)$, there is one location showing a very strong peak, but elsewhere the distribution appears diffuse. Furthermore, the shape during diffusion can be reproduced as either a distorted shape or a clean elliptical shape depending on the combination of standard deviations within p_k.

C. The Effect of the Correlation Coefficient in p_k

Figure 6 investigates the influence of ρ_1 and ρ_2 for each p_k on the distribution of the mixture model. Here, $\mu_{NN_1},\ \mu_{NN_2}$

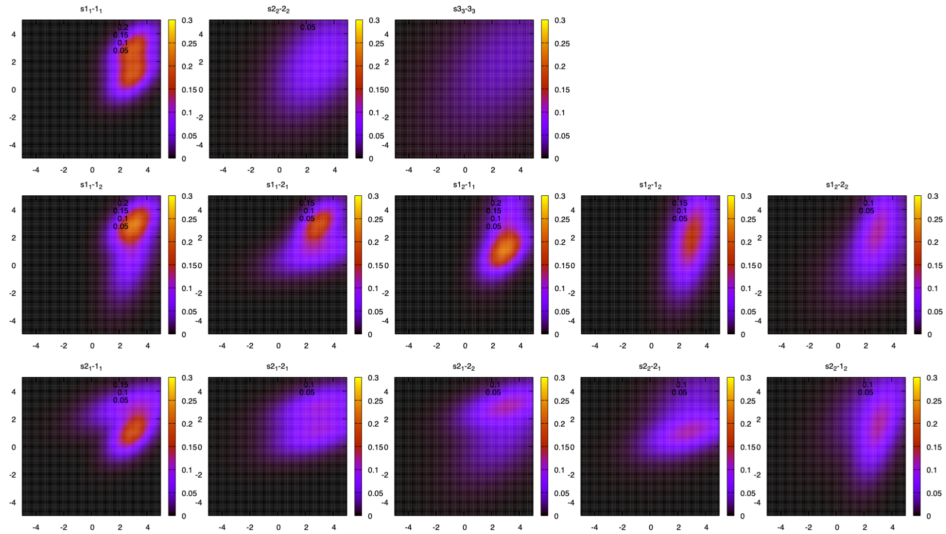

Figure 5. Two dimensional normal distribution for $K = 2$ when fixed at $(\mu_x,\ \mu_y) = (3,\ 3), (3,\ 1), \rho = 0.5$. The figure does not show cases where either σ_{NN_1} or σ_{NN_2} equals 3.

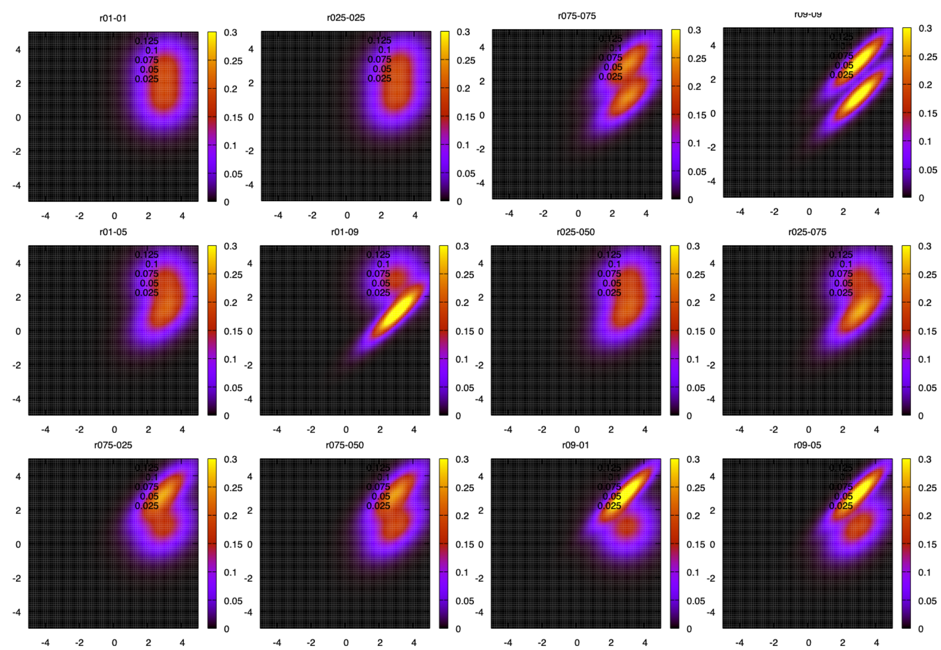

Figure 6. Two dimensional normal distribution with $K = 2$ when $\mu_x, \mu_y = (3, 3)$, $\mu_x, \mu_y = (3, 1)$, and $\sigma_x, \sigma_y = (1, 1)$ are fixed.

for p_1 were set to $(3,\ 3)$, $\mu_{NN_1},\ \mu_{NN_2}$ for p_2 as $(3, 1)$, and $\sigma_{NN_1},\ \sigma_{NN_2}$ were varied from 1 to 3. In the figure, the top row shows the $\rho_1 = \rho_2$, the middle row shows the $\rho_1 < \rho_2$, the bottom row shows the $\rho_1 > \rho_2$.

As a general trend, as the value of ρ increases, p_k shows a concentration of spread in specific directions toward the center, and the probability density takes on sharply large values. That is, it exhibits bias. However, this bias appears to occur generally around $\rho > 0.7$. Regarding shape, when the centers of p_k are close together, the diffusion tendency is not lost even if the values of ρ for each k are small. Directivity becomes pronounced only when large values of ρ are observed on one side exclusively.

D. The Effect of Average and Correlation Coefficient on the Mixture Model in p_k

The figure depicts a mixture model of two dimensional normal distributions, with σ_{NN_1} and σ_{NN_2} each fixed at 1 for p_k. In the figure, The top row fixes $(\mu_{NN_1},\ \mu_{NN_2})$ at $(0.5,\ 3)$, $(0.5,\ \text{-}3)$ The middle row fixes $(\mu_{NN_1},\ \mu_{NN_2})$ at $(1,\ 1)$, $(-1,\ -1)$. The bottom row fixes $(\mu_{NN_1},\ \mu_{NN_2})$ at $(3,\ 3)$, $(3,\ 1)$. Additionally, from the left column to the right column, the values of $(\rho_{p_1},\ \rho_{p_2})$ were varied to $(0.1,\ 0.1)$, $(0.1,\ 0.5)$, $(0.5,\ 0.1)$, $(0.5,\ 0.5)$. The displayed colors represent probability density. As the color transitions from blue to yellow, it indicates a higher probability density for the corresponding $(d_{NN_1},\ d_{NN_2})$, meaning the frequency

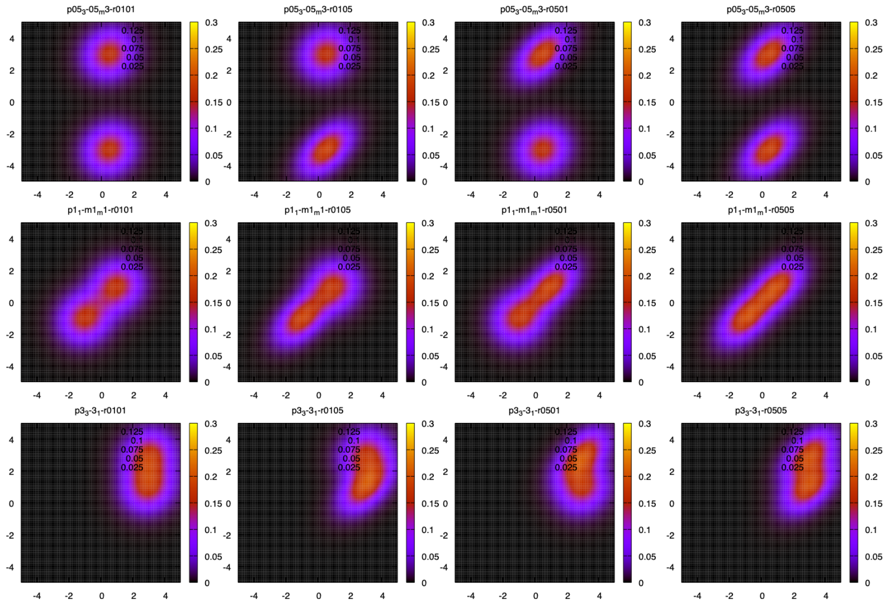

Figure 7. Two dimensional normal distribution for $K = 2$ when fixed at $(\sigma_x,\ \sigma_y) = (1,\ 1)$.

of occurrence for the degree with that value increases.

Comparing the upper and lower parts of the figure, when $\mu_{NN_1},\ \mu_{NN_2}$ are sufficiently far apart, $p_1,\ p_2$ have independent distributions. However, as $p_1,\ p_2$ approach each other, the two distributions are mixture. Specifically, when varying $(\rho_{p_1},\ \rho_{p_2})$ up to $(0.5, 0.5)$, depending on the relative positions of $(\mu_{NN_1},\ \mu_{NN_2})$ between p_k, they may appear to form a single elliptical distribution. Furthermore, comparing the left and right sides, as the values of ρ_{p_1} and ρ_{p_2} increase—that is, as the correlation strengthens—the distributions of both extend along the major axis of the ellipse, distorting their shape. However, no single point exhibits a decisively high density; instead, relatively widespread areas with moderately high values are observed.

E. The meanning of k, μ, σ, ρ in the Depth Structure of Nearest Neighbor

Based on the above, we now discuss the meaning of k, μ, σ, ρ in the proposed depth structure of nearest neighbor.

First, k indicates the number of core probability distributions in the depth structure of the nearest neighbor plane. When all respondents provide the same answer, the expected k would be 1 or a very small value. As responses become more diverse, the expected value of k increases. However, if there is too little consensus among respondents, setting k becomes meaningless, and the probability distribution for the plane will essentially diverge.

$(\mu_{NN_1},\ \mu_{NN_2})$ indicates positions on the same plane. If $(\mu_{NN_1},\ \mu_{NN_2})$ are separated between p_k, we consider the interference between p_k is able to be negligible. Since $(\mu_{NN_1},\ \mu_{NN_2})$ for p_k indicates the properties of connecting words for W_j, the existence of isolated high probability density regions at specific positions implies that the respondent group is using words with similar characteristics. Particularly when the number of k is small, we consider that learners' preferences lack diversity due to specific social factors like lectures or habits. When focusing on short answers to semi-open-ended questions, responses are mostly based on learned concepts, so the words contained in short answers should be somewhat limited. In this sense, it is unlikely that k would be large.

Next, $(\sigma_{NN_1},\ \sigma_{NN_2})$ indicates the degree of diversity within the same plane. The magnitude of σ_{NN_1} for a given p_k is expected to be small, as using uniform words for explanatory concepts likely restricts the number of connectable words. Conversely, the trend for σ_{NN_2} will vary significantly depending on how extensively explanations are provided or whether additional conceptual explanations are given. If additional explanations of the concept are provided before a deep understanding of the learned content is achieved, we consider the σ_{NN_2} will be tend to small. Conversely, when providing additional explanations of the concept, the diversity of words used exists depending on the perspective from which the additional explanation is made. Therefore, we consider the σ_{NN_2} will be tend to large.

Finally, we consider the meaning of ρ_k in the plane. The ordinal sense of a correlation coefficient indicates the strength of the relationship between d_{NN_1} and d_{NN_2}. While d_{NN_1} and d_{NN_2} cannot be expressed as functions describing this relationship, they can be positioned as properties of W_{j_α} that are connectable to W_j. Since most W_j are considered linkable to both global words and specific words, ρ is expected to take very small values. However, for p_k regarding specific words, if only words used exclusively for explaining specialized concepts can be employed, then ρ would likely exhibit strong correlation. In this sense, ρ should be useful for exploring the presence or absence of such special conditions.

From the above, it indicates that when analyzing p_k on the depth structure of the nearest neighbor plane, μ and σ are the most important factors, but occasionally ρ is also necessary. Moving forward, it will be necessary to demonstrate the validity of this series of considerations regarding the parameters when applying them to actual data.

IV. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a framework that classifies learners' short answers into three levels as one method for grading semi-open-ended questions, requiring only minimal knowledge of complex networks. This framework enables the classification of learners' level of understanding and degree of knowledge expression, and contributing to subsequent educational efforts. By adapting the concept of the mixed Gaussian model, we created a depth structure of the nearest neighbor plane for WCN. We demonstrated that by adjusting the number of p_k distributed within it and the effects of variations in μ, σ, and ρ, as well as the synthesis process, it is possible to reproduce various distributions. In practice, obtaining approximations of probability density for this plane it requires preparing real data and deriving approximate probability density using methods like the EM algorithm. Regarding real data, we plan to validate this approach using short answer data from thousands of semi-open-ended questions on graph reading comprehension, for example, obtained by Yagashira et al. [19]. Additionally, a key advantage of this method is its independence from the language used. We aim to apply it to other languages in the future to confirm its usefulness. We believe this method is beneficial when semi-open-ended questions can be posed, assuming there are settings like education or lectures where learners share a common understanding. On the other hand, considering the contexts in which this approach is applied, we assume that the realistic application of this method is for small-scale settings, such as within a classroom or a school. For this reason, while application to large-scale datasets is not currently anticipated. If we consider applying this approach to large-scale datasets in the future, instability—such as an excessive number of degree—may arise, requiring appropriate measures to address it.

REFERENCES

[1] M. Sedighi, “Using of co-word analysis method in mapping of the structure of scientific fields(case study: The field of informetrics)”, *Iranian Journal of Information Processing Management*, vol. 30, pp. 373–396, Dec. 2015.

[2] T. Li, J. Bai, X. Yang, Q. Liu, and Y. Chen, “Co-occurrence network of high-frequency words in the bioinformatics literature: Structural characteristics and evolution”, *Applied Sciences*, vol. 8, no. 10, 2018, ISSN: 2076-3417.

[3] W. Liang, “Spectra of english evolving word co-occurrence networks”, *Physica A: Statistical Mechanics and its Applications*, vol. 468, pp. 802–808, 2017, ISSN: 0378-4371.

[4] M. Garg and M. Kumar, “The structure of word co-occurrence network for microblogs”, *Physica A: Statistical Mechanics and its Applications*, vol. 512, pp. 698–720, 2018, ISSN: 0378-4371.

[5] M. Fudolig, T. Alshaabi, M. Arnold, C. Danforth, and P. Dodds, “Sentiment and structure in word co-occurrence networks on twitter”, *Applied Network Science*, vol. 7, p. 9, Feb. 2022.

[6] D. Amancio, J. Machicao, and L. Quispe, “Leveraging word embeddings to enhance co-occurrence networks: A statistical analysis”, *PLOS One*, vol. 20, Jul. 2025.

[7] T. Millington and S. Luz, “Analysis and classification of word co-occurrence networks from alzheimer's patients and controls”, *Frontiers in Computer Science*, vol. 3-2021, 2021, ISSN: 2624-9898.

[8] K. Zupanc and Z. Bosnić, “Increasing accuracy of automated essay grading by grouping similar graders”, in *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, ser. WIMS '18, Novi Sad, Serbia: Association for Computing Machinery, 2018, pp. 1–6, ISBN: 9781450354899.

[9] T. Wang, N. Inoue, H. Ouchi, T. Mizumoto, and K. Inui, “Inject rubrics into short answer grading system”, in *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, C. Cherry et al., Eds., Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 175–182.

[10] A. Condor, Z. Pardos, and M. Linn, “Representing scoring rubrics as graphs for automatic short answer grading”, in *Artificial Intelligence in Education*, M. M. Rodrigo, N. Matsuda, A. I. Cristea, and V. Dimitrova, Eds., Cham: Springer International Publishing, 2022, pp. 354–365, ISBN: 978-3-031-11644-5.

[11] L. Zhang, Y. Huang, X. Yang, S. Yu, and F. Zhuang, “An automatic short-answer grading model for semi-open-ended questions”, *Interactive Learning Environments*, vol. 30, no. 1, pp. 177–190, 2022.

[12] T. Lertchaturaporn and P. Songmuang, “Automated scoring model for semi-open-ended question in scientific explanation”, in *2024 IEEE International Conference on Cybernetics and Innovations (ICCI)*, 2024, pp. 1–5. DOI: 10.1109/ICCI60780.2024.10532484.

[13] R. Conijn, P. Kahr, and C. Snijders, “The effects of explanations in automated essay scoring systems on student trust and motivation”, *Journal of Learning Analytics*, vol. 10, no. 1, pp. 37–53, Mar. 2023.

[14] F. Jamil and I. A. Hameed, “Toward intelligent open-ended questions evaluation based on predictive optimization”, *Expert Systems with Applications*, vol. 231, p. 120 640, 2023, ISSN: 0957-4174.

[15] L.-H. Chang and F. Ginter, “Automatic short answer grading for finnish with chatgpt”, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, pp. 23 173–23 181, Mar. 2024.

[16] C. Grévisse, “Llm-based automatic short answer grading in undergraduate medical education”, *BMC Medical Education*, vol. 24, Sep. 2024.

[17] R. Ferreira Mello et al., “Automatic short answer grading in the llm era: Does gpt-4 with prompt engineering beat traditional models?”, in *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, ser. LAK '25, Association for Computing Machinery, 2025, pp. 93–103, ISBN: 9798400707018.

[18] K. T. Nakahira, M. Watanabe, and M. Kitajima, “Assessment of developmental stages of generic skills: A case study”, in *Proceeding of the 22nd International Conference on Computers in Education, ICCE 2014, Nara, Japan, November 30 - December 4, 2024*, Asia-Pacific Society for Computers in Education, 2014, pp. 200–205.

[19] Y. Yagashira, K. T. Nakahira, M. Arai, G. Kumoi, and T. Yukawa, “Proposal of a method for estimating the acquisition of graph comprehension ability based on students' descriptive answers”, *Procedia Computer Science*, vol. 270, pp. 1856–1865, 2025, 29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025), ISSN: 1877-0509.