IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 January 2023, accepted 4 February 2023, date of publication 13 February 2023, date of current version 16 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244614

APPLIED RESEARCH

Optimal Knowledge Component Extracting Model for Knowledge-Concept Graph Completion in Education

HYUNHEE CHOI[1], HAYUN LEE[1,2], AND MINJEONG LEE[1,3,4]
[1]LAIVDATA, Seoul 06633, South Korea
[2]School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea
[3]Da Vinci College of General Education, Chung-Ang University, Seoul 06974, South Korea
[4]Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul 02841, South Korea

Corresponding author: Minjeong Lee (minjeonglee@cau.ac.kr)

ABSTRACT As people have become accustomed to non-face-to-face education because of the COVID-19 pandemic, adaptive and personalized learning is being emphasized in the field of education. Learning paths suitable for each student may differ from those normally provided by teachers. To support coaching based on the concept of adaptive learning, the first step is to discover the relationships among the concepts in the curriculum provided in the form of a knowledge graph. In this study, feature reduction for the target knowledge-concept was first performed using Elastic Net and Random Forest algorithms, which are known to have the best performance in machine learning. Deep knowledge tracing (DKT) in the form of a dual-net, which is more efficient because of the already slimmer data, was then applied to increase the accuracy of feature selection. The new approach, termed the optimal knowledge component extracting (OKCE) model, was proven to be superior to a feature reduction approach using only Elastic Net and Random Forest using both open and commercial datasets. Finally, the OKCE model showed a meaningful knowledge-concept graph that could help teachers in adaptive and personalized learning.

INDEX TERMS Deep learning based knowledge tracing (DKT), dual-net, elastic net, feature selection, knowledge component (KC), least absolute shrinkage and selection operator (LASSO), random forest (RF).

I. INTRODUCTION

As travel and social contact have been restricted because of the COVID-19 pandemic, not only people's daily activities but also the way they work and communicate in society are rapidly shifting toward a non-face-to-face modality [1]. Inevitably, the field of education has also rapidly adapted to an Internet-based learning management system (LMS), which can guide and evaluate student learning online. Prior to this, the digital transformation of education using online education systems was already underway, for instance, through open online education platforms such as massive open online courses (MOOCs) that allow students to take courses according to their choice, regardless of time and location [2]. In online learning environments, including MOOC platforms, personalization of learning with computing technology may be considered [3], [4].

Personalized learning is a method of suggesting a learning speed and process customized according to the characteristics of each learner's prior experience, knowledge, and learning pattern; that is, adaptive learning that uses big data and artificial intelligence technology to recommend the optimal learning path or provide feedback based on individual learning patterns and results by analyzing a large amount of educational data in real time. A high level of personalized learning can be implemented using adaptive learning techniques [5], [6]. The knowledge that learners are required to acquire in a curriculum can be expressed in a knowledge-concept graph that consists of key concepts, such as knowledge components (KCs) and their relationships. Learners' knowledge tracing or learner-specific learning paths can be recommended based on a knowledge-concept graph [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

In general, the structure of knowledge and the learning path of a subject are defined by the instructor in the curriculum. However, relationships with other concepts that affect a specific learner's acquisition of a certain knowledge-concept might differ from the structure of the curriculum or show a hidden aspect. Therefore, if we analyze correlations among KCs based on the acquisition status of individual KCs corresponding to the knowledge concept structure of the curriculum and discover neighboring KCs that have a significant influence on a given KC, we can derive a practical KC graph that complements the given curriculum and provides precise and effective coaching to learners.

Bayesian knowledge tracing (BKT), a well-known technique for tracing learners' knowledge, uses Bayesian probability estimation which leaves learners' question responses in two states: acquired and non-acquired [7]. BKT has the limitation of matching the learner's knowledge status to only two cases based on the response to a question. To overcome this, DKT technique has emerged, which traces a learner's knowledge by applying deep learning based on recurrent neural networks (RNN) [8], which further reflects the precedence of knowledge elements or context [9]. Methods that extend DKT to apply phase prediction [10] and difficulty levels [11] have been proposed. If it is possible to predict the learning state of a concept using the acquisition state of concepts that are highly related to it in deep learning, performance improvement in the knowledge tracing model can be expected. Therefore, it is necessary to introduce a feature selection technique to apply the deep learning–based algorithm used in the knowledge tracing model in an efficient and explanatory way [12], [13], [14], [15]. As a feature selection algorithm for deep learning, Ambroise and McLachlan [15] devised a dual-net architecture comprising a selector net that selects major features based on deep learning performance evaluation and an operator net that evaluates performance by learning the selected features through deep learning interaction.

In this study, the optimal knowledge component extraction (OKCE) model is devised to complete the substantial knowledge-concept graph by selecting valid and optimized neighbor knowledge components. The OKCE model uses the Elastic Net with LASSO parameters (hereinafter referred to as Elastic Net (LASSO)) and the Random Forest (RF) algorithm to select neighbor KCs that are highly correlated with each KC from the significant KC pairs. Then, the OKCE model uses the dual-net-based feature selection method to obtain the optimal neighbor KCs that have an increased impact on the learning prediction accuracy. As one of the components of the dual-net, the operator net calculates the influence of the neighboring KCs on the learning state of each KC using DKT. The other component of the dual net, the selector net, extracts the optimal neighbor KC subset based on the feedback from the operator net. To verify the OKCE model, experiments were performed on a dataset released at the 2010 KDD Cup competition and a dataset of k6-level mathematics by an educational commercial service in South Korea.

This paper makes the following main contributions:

- The optimal KC subset extracted by the OKCE model can complete the KC graph and support finding a meaningful knowledge-concept path for adaptive and personalized learning.
- The optimal KC subset can show higher DKT performance on a target KC, even when the number of KCs is limited, because the OKCE model prudently extracts the important KCs using the dual-net among the highly correlated relations obtained through the Elastic Net (LASSO) + RF method.

The remainder of this paper is organized as follows. Section II examines the characteristics of algorithms that derive KC relationships, creates knowledge maps that reflect the characteristics of the educational field, and reviews existing research on feature selection to screen for KCs that have an important impact on learning analysis. Section III introduces the structure and experimental data of the proposed OKCE model to derive a KC subset that constitutes a knowledge map with a high correlation for each KC. Section IV describes the operating process of the OKCE model in detail. The results of applying the model to derive the KC relationship proposed in this study to actual KDD Cup data and K6 mathematics level assessment data are analyzed in Section V. Finally, Section VI concludes the paper.

II. RELATED WORKS

A. FINDING RELATIONS

Associations among concepts can be obtained by performing association rule mining (ARM) [16] or by analyzing correlations between variables. Chen et al. [9] processed various types of data into transaction forms using neural sequence labeling and then performed ARM [16] using KC data. Based on this, an attempt was made to find a relational map of KCs. Son et al. [17] proposed a learning path recommendation model that considered learners' background knowledge and goals. This metaheuristic model was developed using a genetic algorithm (GA) and ant colony optimization (ACO) algorithm. Lu et al. [18] proposed a model to trace students' knowledge status using a recurrent neural network (RNN) [19] that exploits the relationship and topology information among KCs.

On the other hand, the concept of link prediction offers a method of discovering the relationship between nodes using the feature data of the node rather than the test data. One option is to use the tensor decomposition method and domain information of entity type from an early model, as in Chang et al. [20], and the neural network–based link prediction model of Dettmers et al. [21], as in a recent study. Dettmers et al. presented a case in which performance was improved by performing a simple transformation on convolutional 2D knowledge graph embeddings (ConvE) based on a convolutional neural network (CNN). The hypernetwork knowledge graph embeddings (HypER) used in [22] provide an example of using a simple calculation method to solve sparsity, as well as a mechanism for sharing weights.

A characteristic of prior studies is that in identifying meaningful relationships, they moved away from measuring simple correlations and attempted to solve them in the form of supervised learning. That is, as shown in ARM [16], the strength of the association between individual KCs can be determined using characteristic values such as confidence [23]. However, if the concept of supervised learning is used, it is possible to compare it with other relations [24], and their roles as the basis for using a regularization model [25] related to regression coefficients can be obtained.

B. FEATURE SELECTION

Many different features of educational data can be identified by using the KC approach. If the process of finding relationships among concepts is carried out without feature selection, correlations among features inevitably lead to the creation of models with less explanatory power, as identified in the Curse of Dimensionality [26]. Thus, extracting and identifying an effective feature set is the key to this research.

Feature selection is generally divided into filter, wrapper, and embedded methods [12]. The wrapper method includes simulated annealing (SA) [13], which mimics the quenching process of metals, in addition to the traditional forward, backward, and sequential selection methods and genetic algorithms (GA) [14] used to find combinations of features that predict optimal responses [14]. Ambroise and McLachlan [15] noted that in some cases, the uncertainty arising from feature selection may be much greater than the uncertainty inherent in the post-selection model; therefore, selection bias may appear during feature selection, which may adversely affect the performance of the model. Kuhn and Johnson [12] argued that, if the training dataset is small, proper resampling is essential before performing feature selection.

1) FEATURE SELECTION ALGORITHMS

Ploeg and Steyerberg [27] used the nested bootstrap technique [28] to develop feature selection and predictive models for DNA microarrays of 222 Legionella pneumococcal strains to determine the performance of the classification and regression trees (CART) [29], random forest (RF) [30], support vector machine (SVM) [31], and LASSO [32] methods in terms of AUC [33]. We compared these methods using the AUC as a criterion. The AUC values for CART, RF, SVM, and LASSO were 0.937, 0.938, 0.887, and 0.965, respectively. RF showed the highest AUC value and excellent performance. Lu and Petkova [34] performed feature selection to select items to be analyzed from among questionnaire items to develop tools for screening mental illness. LASSO, Elastic Net [25], CART, RF, and two-sample t-tests were considered as candidates for feature selection algorithms, of which LASSO and Elastic Net showed excellent performance.

Zou and Hastie [25] explained that when highly correlated explanatory variables exist, LASSO randomly selects one of these variables to reduce the coefficient, whereas Elastic Net groups the highly correlated explanatory variables and selects or removes them all. Therefore, if there is a high correlation among multiple explanatory variables, Elastic Net rather than LASSO is recommended for feature selection.

2) FEATURE IMPORTANCE RANKING FOR DEEP LEARNING

In deep learning, feature importance ranking (FIR) measures the contribution of individual features (variables) to supervised learning performance [35]. FIR is closely related to the aforementioned feature selection because it calculates the feature importance from the optimal subset and can be used as a proxy for feature selection [35], [36]. Wojta and Chen [35] noted that in machine learning, LASSO and RF are known to achieve the highest level of FIR but have limited learning capabilities compared to deep learning and will not always act on complex dependencies between input values and target variables.

3) DUAL-NET ARCHITECTURE FOR DEEP LEARNING

Yang et al. [37] proposed a dual-net architecture consisting of two steps: selecting a feature and then learning the selected feature to optimize deep learning. The dual-net architecture considers model quality and training efficiency and is characterized by high learning accuracy and low computational cost.

Wojta and Chen [35] configured a dual-net comprising a selector net and operator net. The selector net selects features that affect the performance of deep learning from a given feature set and passes them to the operator net, which applies a CNN (a deep learning technique) with the selected features to calculate the accuracy, and then passes them back to the selector net.

Sundara et al. [38] introduced a dual deep learning architecture (DDLA): a network consisting of a feature extractor (MobileNet) and a classifier (DenseNet-121). This architecture performs deep learning after machine learning (ML). Standard (Flavia, Folio, and Swedish Leaf) and custom collection (Leaf-12) data were used as inputs, and the DDLA was used to assess the data with 98.71%, 96.38%, 99.41%, and 99.39% accuracy, respectively. The experiment also showed superior results in terms of computation time compared with a single deep learning architecture.

C. KNOWLEDGE TRACING

Knowledge tracing (KT), which originated around the time of Corbett and Anderson [7], traces and predicts students' changing learning status. Knowledge tracing techniques include BKT, probabilistic models of dynamic Bayesian knowledge tracing (DBKT) [39], logistic models such as performance factor analysis (PFA) [40], and deep learning models such as DKT [8] with nonlinear model functionality.

In general, the test data for a question included information about the time the student responded together with KC data related to the question. For learning data that includes the dimension of time, there is a trend to apply RNNs [40], [41] that model how students' knowledge levels change over time

TABLE 1. Four datasets with learning logs.

Dataset	Number of logs	Number of students	Number of KCs	K-12 grade
KDD_RATIO	179,001	1,146	19	6~7
SSM_11	419,526	4,971	36	5~6
SSM_12	415,020	3,892	31	6~7
SSM_13	240,240	2,539	22	6~7

KDD_RATIO represents the dataset configuration for the ratio-related KDD Cup data. SSM_11–13 in Table 1 refer to the Company D dataset configurations for levels 11–13. Column "K-12 grade" represents learners' educational background.

to the concept of basic DKT, which is also characterized as long short-term memory (LSTM). LSTM is in the RNN family but compensates for the shortcomings of the vanishing gradient approach [42].

On the other hand, Şahín and Diri [43] showed that accuracy of feature selection can be further improved by applying regularization and variable selection using Elastic Net as verified by Zou and Hastie [25] for the DKT model, with LSTM [44], [45] then showing strong performance in time series data.

III. PROPOSED METHOD

A. PURPOSE OF THE STUDY

The purpose of this study is to obtain the optimal subset of KC pairs from the significant relationships among KCs, considered as knowledge graph nodes, using feature selection techniques. To this end, we propose an OKCE model that performs feature selection or FIR using a deep learning–based dual-net architecture after reducing the number of KCs that are targeted for feature selection using machine learning algorithms, such as Elastic Net. In the feature selection step, DKT applied with LSTM is selected for the operator net of the dual-net architecture based on the time-series properties of the training data, thereby improving the accuracy of the KC relationships extracted through the feature selection process.

To verify that the proposed OKCE model improved the performance in detecting the optimal subset of KC pairs compared to the Elastic Net (LASSO) + RF model, the subset of KC pairs obtained by Elastic Net (LASSO) + RF was used as an experimental control group. Performance was analyzed by comparing the measures of the control and OKCE groups.

B. OVERVIEW OF DATASET

To demonstrate the explanatory power of the research model, the 2020 K-12 Math evaluation data of Korean Company D, where the KC relationship target exists, and the algebraic practice data (hierarchy = 'Unit RATIO-PROPORTION, Section RATIO-PROPORTION-3', hereinafter referred to as "ratio" data) from the 2010 KDD Cup data mining contest were used to test the efficiency of the algorithm. Table 1 presents the data statistics.

The learning accuracy for each KC in these datasets was defined as the average value of the responses, measured as 0 (unlearned) and 1 (learned) for the questions related to each KC, as shown in Fig. 1.

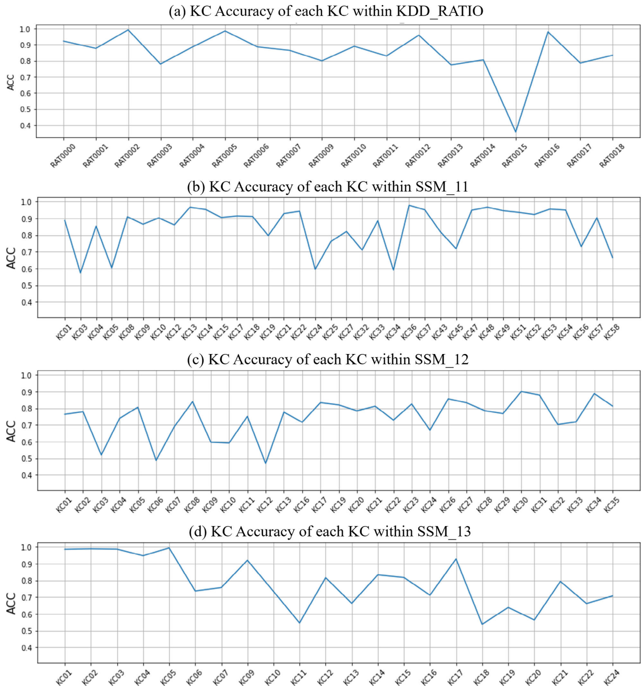


FIGURE 1. Learning accuracy graph for each dataset for the experiments.

The four datasets, KDD_RATIO and SSM_11-13, used for the OKCE model were log data generated during the students' problem-solving attempts. The main pre-processing step is to transform the log data in the form of transactions into a matrix in which the KC names for each learner, which are the target and feature variable names, are stretched horizontally. Additional pre-processing steps are based on the "garbage in, garbage out" principle [46].

The process is as follows:

- Delete records with all correct or incorrect answers for all KCs.
- Remove variables with all correct or incorrect answers for all learners.

C. STRUCTURE OF OKCE MODEL

The OKCE model is designed as a feature selection step to list KC variables that are highly correlated with a target KC variable by applying the Elastic Net (LASSO) + RF algorithm and a dual-net consisting of an operator net using LSTM and a selector net that selects the optimal feature set based on the score of the operator net. The structure of the proposed approach is shown in Fig. 2.

By applying the Elastic Net (LASSO) + RF algorithm, m KCs that show a high correlation with the learning state of a target KC are extracted as a candidate KC subset, where m is set between one-half and one-third of the total number of KCs in the dataset. In this study, m is set to 10 considering the numbers of KCs in the four datasets. Next, an optimal KC

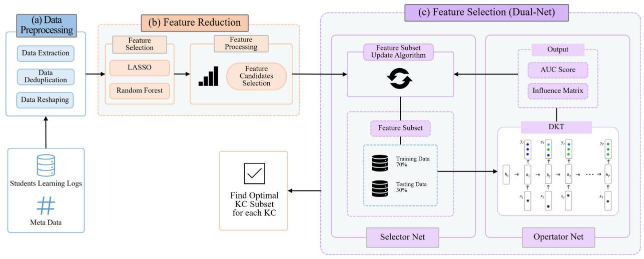


FIGURE 2. System structure and operating flow diagram of OKCE model. (a) In the data pre-processing step, the training data for each question is transformed into training data for each KC. (b) In the feature reduction step, a meaningful relationship with each KC is extracted by the Elastic Net (LASSO) + RF algorithm. (c) In the dual-net structure utilizing DKT in the feature selection step, each KC derives an optimal KC subset that maintains the performance of DKT at a high level among neighboring KCs with a significant relationship derived from (b).

subset consisting of n (=m/2) more important KCs is derived through dual-net from the candidate KC subset.

IV. PROCEDURE OF THE PROPOSED METHOD

This section introduces the workflow of the OKCE model when applying the Elastic Net (LASSO) + RF algorithm to select significantly related KCs for each KC in the dataset, and applying dual-net to extract the optimal meaningful KC subset from the significant KC subsets derived in the previous step.

A. FEATURE REDUCTION BY ELASTIC NET (LASSO)+RF

This study extends the task of finding the correlation between KCs and the problem of using regression analysis to find related KC pairs based on the mathematical association between correlation coefficients and regression coefficients [47]. We applied multivariate regression analysis to predict a target variable more accurately. KC pairs (a target variable and a related explanatory variable) can be selected through Elastic Net (LASSO), which simultaneously considers multicollinearity [48] and ensemble theory at the same time. The objective function for determining the regression coefficients of the Elastic Net (LASSO) is shown in (1). Let

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} 1 & X_1^{(1)} & X_2^{(1)} & \cdots & X_p^{(1)} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & X_1^{(N)} & X_2^{(N)} & \cdots & X_p^{(N)} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix};$$

Then

$$\hat{\beta} = argmin\,(y - X\beta)^T\,(y - X\beta)\,, |\beta| < t. \tag{1}$$

In (1), β is a regression coefficient and t is a free parameter that determines the strength of the regularization. However, although the significance level is limited to the extraction of meaningful relationships, the number of relationships that pass a certain level may be too large. In this study, a random forest with the function of ordering the importance of features as a nonlinear model was applied. The importance of the features is calculated as in (2) [49], [50].

$$VI^{(t)}(X_k) = \sum_{i\in\beta^{(t)}} \frac{\left(\hat{y}_i^{(t)} - y_i\right)^2}{\left|\beta^{(t)}\right|} - \sum_{i\in\beta^{(t)}} \frac{\left(\hat{y_{i(k)}}^{(t)} - y_i\right)^2}{\left|\beta^{(t)}\right|} \tag{2}$$

where

- $i = 1, \ldots, N (N =$ number of observations),
- $\beta^{(t)}$ means a tree based on out of bag (OOB) sample,
- X_k is the k^{th} feature (explanatory) variable of the model, that is, the k^{th} KC variable,
- y_i is the predicted value of the i^{th} observation of the target variable,
- $y_{i(k)}$ is the predicted value when permuted with respect to X_k, and
- t is the tree generated by the RF algorithm.

In brief, Elastic Net (LASSO) extracts KCs that are unlikely to correlate with different KCs among the input (i.e., feature) variable KCs. Regression with a target KC is then performed to extract KC pairs that satisfy the 10% significance level. Next, the RF algorithm was applied to calculate the importance of the input KCs. If the importance is within the top 20%, select KC, which means considering the pair (target and feature variables) as a significant pair.

In general, 5% is used as the significance level for many statistical tests, but 10% is considered as the significance level of Elastic Net (LASSO) for experiments on the performance of the dual-net architecture and is the accepted significance level for social science data [51], [52], [53]. Additionally, in the Random Forest, we selected the top 20% pairs showing optimal overlap with the number of relations emerging from the Elastic Net (LASSO). Among the KC pairs derived from the Elastic Net (LASSO) and RF algorithms, more meaningful relationship pairs should be selected for efficient feature selection.

The criteria for selecting meaningful pairs are as follows:

- assign "BOTH" if a pair is found to be significant in both Elastic Net (LASSO) and Random Forest
- assign "LASSO" if a pair satisfies the 10% significance level.: Among the relationship pairs satisfying the 10% significance level, if a relationship pair does not match criterion 1, name it LASSO and select it as a valid relationship pair.
- assign "RF" if a pair satisfies 20% importance: Among the relationship pairs satisfying the 20% level of importance, a relationship pair not satisfying criteria 1 and 2 is named RF and selected as a valid relationship pair.
- select random pair if insufficient: If the number of pairs selected through criteria 1–3 does not reach the target number of relationships, any random pair is selected from among the pairs not selected in the significance level or importance criteria.

The above criteria constitute a procedure for applying Elastic Net (LASSO) and RF to the preprocessed learner data

(a) Pseudocode for Feature Selection Procedure using Dual-Net

Algorithm 1

```
Input: Preprocessed educational data
Output: Optimal KC subset for each KC

/*feature reduction */
Function SelPartialKCs(processed data)
   target_kcs, kc_candidates ← Feature selection using ElasticNet(LASSO) and Random Forest
   return target_kc, kc_candidates

/* operator net : find optimal KC subset for each KC */
Function DoOperatorNet(model, now_ kcs)
   model.train(now_ kcs)
   calculate influence_matrix
   return before_ kcs, cur_AUC

/* selector net : find new KC subset and update best_AUC, best_kcs, before_kcs */
Function JudgeAUC(cur_AUC, before_ kcs)
   new_kcs ← Judgment Algorithm
   return new_kcs


kcs, kc_candidates ← SelPartialKCs (processed data)

for each (target_kc, target_kc_cands) ∈ (kcs, kc_candidates) do
   /* initialize the operator net */
   model ← OperatorNet.init()

   /* setup the variables from initial operation with top 5 target_kc_cands */
   best_kcs ← select top 5 KCs of target_kc_cands
   now_kcs, best_AUC ← DoOperatorNet(model, best_kcs)

   /* prepare the second kc subset and its AUC in order to compare with the initial subset */
   now_kcs ← randomly select 2 KCs from top 5 and  3 KCs from bottom 5 of target_kc_cands
   before_kc, cur_AUC ← DoOperatorNet(model, now_ kcs)

   /* do iteration of selector net and operator net in order to get the optimal KC subset for target KC*/
   for i = 0 to 100 do
      if i == 100 or cnt_best_kcs > 3 do
         break
      now_kcs ←  JudgeAUC(cur_AUC, before_ kcs)
      before_kcs, cur_AUC ← DoOperatorNet(model, now_ kcs)
```

(b) Flowchart for Feature Selection Procedure using Dual-Net

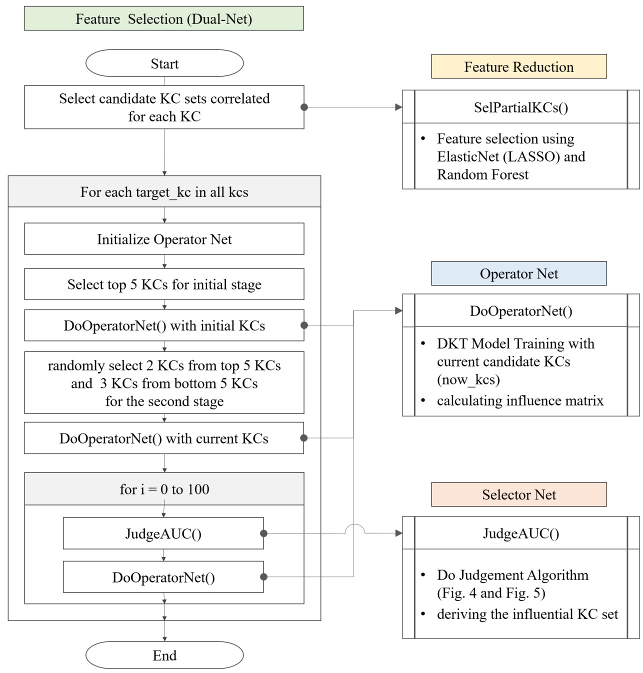


FIGURE 3. **(a) Pseudocode representing the dual-net operating procedure for the OKCE model. (1) Get m relations selected by Elastic Net (LASSO) + RF algorithm. (2) Interact with the operator net based on the DKT learning performance by featuring n neighbor KCs for all candidate KCs. (3) Repeat this operation for each target KC. Repeat (1)–(3) for 100 epochs. However, if the same KC subset is selected more than three times, consider an optimal KC subset to be found and perform an early halt. (b) Flowchart presenting the Algorithm 1.**

to derive KC pairs and finally select m meaningful relations among them with priority. Since the procedure was introduced only for one target KC and the data used in this project included several KCs, the procedure should be repeated for all possible target KCs.

B. FEATURE SELECTION BY DUAL-NET

Among the m important KCs, we aim to identify n (< m) core KCs that have an increased impact on the learning state of the target KC. It can be assumed that the related KCs in the relationship that have a high influence on the learning status of the target KC predict its learning state more accurately. Therefore, the problem of selecting core KCs for a certain KC is the same as that of selecting neighbor KCs that have a significant impact on a certain KC through the feature selection of a deep learning model that predicts the learning state of a target KC.

1) OPERATOR NET

The DKT model [8] is embedded in a net that predicts which questions will be answered correctly, based on the learner's previous learning status. The DKT applied to the operator net was LSTM. It receives the KC subset selected through the selector net and outputs influence scores for the individual KCs of the KC subset.

2) SELECTOR NET

The selector net interacts with the operator net to select key KCs when predicting the learning state of a target KC. To this end, the selector net updates the KC subset according to the frequency history of the selection and the AUC scores, which are the performance indicators returned from the operator net.

As shown in Fig. 3, the dual-net built with a cyclic structure of operator and selector nets finds n (=m/2) optimal KC subsets that have a more significant impact than the m significant KCs derived from the Elastic Net (LASSO) + RF model. More precisely, the initial KC subset of the selector net is determined by the top n KCs extracted from the Elastic Net (LASSO) + RF, based on importance. Thereafter, the next KC subset is updated using the judgment algorithm defined in this study, according to the DKT performance evaluation results of the operator net. The judgment algorithm compares the AUC returned by the operator net with the best AUC to form a new KC subset in the following three cases:

Case 1 When the Current AUC is Higher Than the Best AUC: If the current AUC is greater than the best AUC, it can be assumed that an element in the current KC subset that is not included in the best KC subset plays a role in increasing AUC. Therefore, the difference set between *best_kcs* and *now_kcs* (the current KC subset) is obtained and included in the subset *new_kcs* (the KC subset for the next iteration). Now, we must select the required number of KCs from the candidate KCs to complete the n elements that fulfill the target size of the KC subset. Let k be the number of KCs replenished in this manner.

$$\begin{aligned} new\_kcs &= now\_kcs \cap best\_kcs^{c}, \\ k &= n - \text{size}(new\_kcs). \end{aligned} \tag{3}$$

Case 2 When the Current AUC is not Higher Than the Best AUC but is Higher Than the AUC of the Previous Iteration: The same criteria as in Case 1 are applied to *before_kcs*, the KC subset of the previous iteration, to determine the KC subset for the next iteration:

$$\begin{aligned} new\_kcs &= now\_kcs \cap before\_kcs^{c}, \\ k &= n - \text{size}(new\_kcs). \end{aligned} \tag{4}$$

Case 3 When the Current AUC is Lower Than the AUC of the Previous Iteration: If the current KC set does not have a significant impact on the target KC's learning prediction compared to the best set, ignore the current iteration and restore the picking table to its previous state. The KC subset has been completely reconstructed. That is, we drop all the *now_kcs* and set n to the number of new KCs to be selected.

$$new\_kcs = \{\text{empty}\}, k = n. \tag{5}$$

If k is greater than 0, refer to the KC picking table to select k KCs from the candidate KCs and merge them into *new_kcs*. The criteria for selecting KCs were as follows:

If k is equal to one:
  select one KC randomly from two KCs having upper-bound pick counts
else if k is larger than one:
  select k− 1 KCs randomly from k KCs having upper-bound pick counts
  select one KC randomly from two KCs having lower-bound pick counts

If *new_kcs* configured in this manner are the same as *before_kcs*, perform the KC selection process for *new_kcs* again. When *new_kcs* is completed, the pick value of the KCs included in *new_kcs* increases and is reflected in the selection frequency history. Fig. 4 shows the process of constructing *new_kcs* for Cases 1 and 2. Fig. 5 illustrates the process of Case 3.

When the selection of a new KC subset is completed through the selector net judgment algorithm, the operator net is applied again. The dual net repeats the pairing of the operator and selector nets 100 times for all target KCs, and defines the best KC subset as the optimal KC subset with a high influence on the learning state of the corresponding KC. If the same KC subset was selected three times as the best AUC before 100 iterations, it was regarded as the optimal KC subset, and the feature selection operation of the dual-net was stopped early. It is assumed that the number of features selected by the dual-net, that is, the size n of the KC subset, is greater than three. If index i is the step number out of 100 steps (i.e., epochs), and the number k_i of KCs to be newly selected in the i-th step is 1 while the dual-net is operating, then the number of KCs to be newly selected in the $(i+1)$-th step, k_{i+1}, is always greater than 1:

$$\begin{aligned} k_{i+1} &= n - k_i \text{ for case 1 and case 2}, \\ k_{i+1} &= n \text{ for case 3}. \end{aligned} \tag{6}$$

Therefore, k_i cannot always be 1, and even if k_i alternates between 1 and $n-1$, it covers all i candidate KCs after at most $2n$ iterations. Similarly, as the iteration of the dual-net progresses, the KCs in the subgroup continue to be shuffled; thus, no starvation occurs.

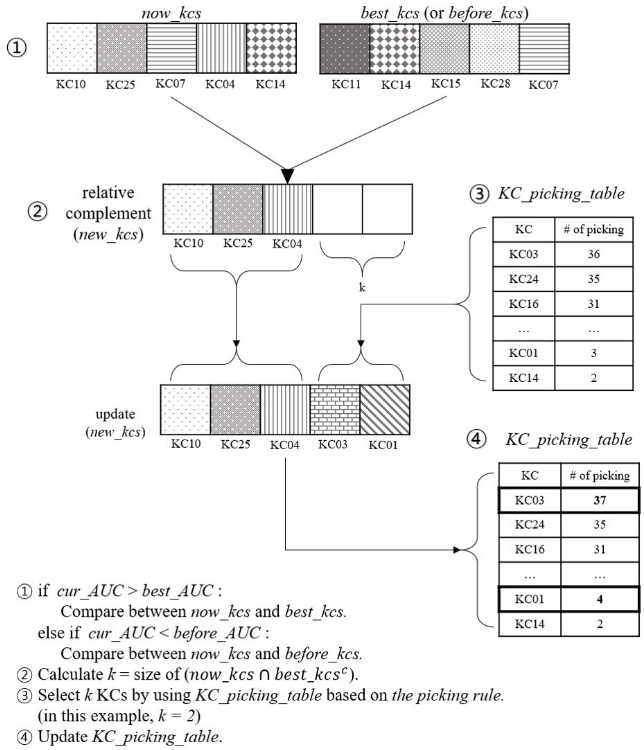


FIGURE 4. Judgment algorithm of selector net for case 1 and case 2: current AUC score is better than best AUC or previous AUC and is not included in the best KC subset or previous KC subset of the current KC subset.

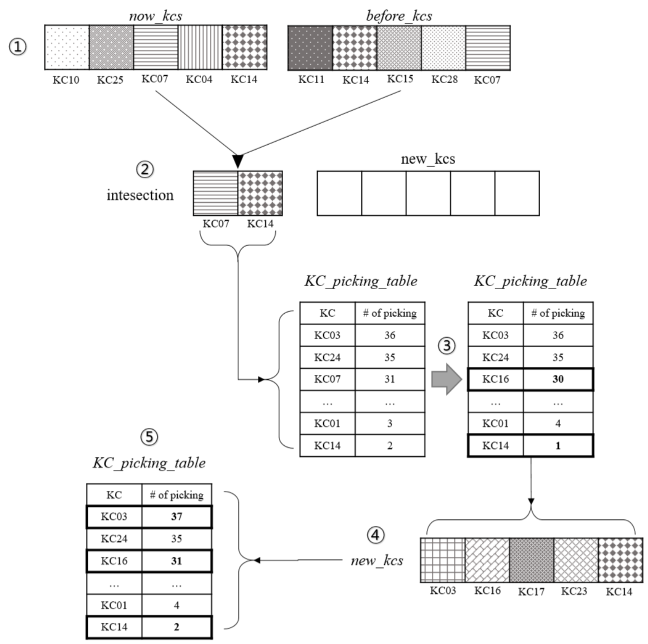
In addition, the judgment algorithm recognizes the effect of new KCs that are not part of the previous *before_kcs* but are newly recruited to *now_kcs* in cases 1 and 2, where the performance of the current operator net has been improved. On the other hand, in case 3, where the performance is worse than before, not only is the current KC subset reset, but also the pick value of the previously added KC is restored, so the adverse effects of learning performance are not reflected in the pick value of the target KC. Therefore, the KC pick value can be considered a significant positive weight for the learning prediction performance of a target KC, and the judgment algorithm can obtain an optimal KC set that is close to the best KC set with the best learning prediction performance of a target KC.

V. EXPERIMENTS

In this section, the OKCE model is used to select the significant KC pairs and to demonstrate that the proposed method improves modeling accuracy using commercial and open data.

A. EXTRACTION OF SIGNIFICANT KC PAIRS BY ELASTIC NET (LASSO) + RF

As the first step of the OKCE model, lists of m significant KC pairs for each KC were derived using the Elastic Net



① Compare between *now_kcs* and *before_kcs*.
② Calculate *now_kcs* ∩ *before_kcs* and initialize *new_kcs*
③ Update the *KC_picking_table*.
④ Randomly select 4 kc of the top-5 and 1 kc of bottom-2
⑤ Update the *KC_picking_table*.

FIGURE 5. Judgment algorithm of selector net for case 3: the current AUC score is worse than the previous AUC score and thus the current KC set is not conducive to prediction.

TABLE 2. Significant KC pairs for KC01 and KC02 of SSM-12, obtained by Elastic Net (Lasso) + RF method.

No.	Before	After	Ord	Method	Imp
1	KC02	KC01	1	BOTH	123
2	KC03	KC01	1	BOTH	51
3	KC31	KC01	2	LASSO	22
4	KC05	KC01	2	LASSO	21
5	KC23	KC01	4	RF	7
6	KC26	KC01	4	RF	19
7	KC24	KC01	4	RF	37
8	KC29	KC01	4	RF	20
9	KC21	KC01	4	RF	15
10	KC08	KC01	4	RF	21
11	KC01	KC02	1	BOTH	150
12	KC03	KC02	1	BOTH	67
13	KC35	KC02	2	LASSO	29
14	KC05	KC02	2	LASSO	20
15	KC08	KC02	2	LASSO	20
16	KC24	KC02	3	RF	40
17	KC27	KC02	4	RF	22
18	KC04	KC02	4	RF	20
19	KC07	KC02	4	RF	13
20	KC16	KC02	4	RF	15

For each KC relationship pair, "Before" and "After" are KC identifiers and indicate feature variables and target variables respectively; the "Ord" and "Method" columns indicate selection criteria to extract input relations for the dual-net at the next step; the "Imp" column is the importance value calculated through the RF algorithm for each pair.

(LASSO) + RF method. Table 2 shows 10 significant KC pairs for KC01 and KC02 for SSM_12, with the selection criteria and importance value for each relation. Table 3 presents the concept titles of each KC within SSM_12.

B. EXTRACTION OF OPTIMAL KC PAIRS BY DUAL-NET USING DKT

The selector net of the dual-net in the OKCE model extracts an optimal KC subset for each KC by evaluating the AUC score returned from the operator net and referring to the pick -count history for each KC. In this process, the requirement is to determine the appropriate size n of the KC subset that possibly shows the effective performance of DKT in the operator net. As shown in Table 4, the highest AUC was observed for $n = 5$. Therefore, in this study, the size of the KC subset selected by the selector net was set to five.

To evaluate the accuracy of the OKCE method, the performance scores of the DKT models trained with each KC subset obtained using different KC selection algorithms were compared. As a comparison group, the DKT model with the top five pairs selected with Elastic Net (LASSO) + RF only, the DKT model with the entire set of KCs, and the DKT model with the best subset among all combinations of KC subsets were generated. For the four datasets, the overall average DKT AUC values of the OKCE model and those of the other three models were compared.

Table 5 shows that the AUC average of the OKCE model is superior to the AUC average of the Elastic Net (LASSO) + RF–only model for all datasets and is close to the best AUC average. Therefore, the selector net of the OKCE model's dual-net can find stronger positive relationships than the other models despite the limited number of KCs.

To take a closer look, we unpacked the DKT performance for KC_i while various KC extraction models were applied. Fig. 6(a) shows the performance of the DKT model using a significant KC subset of size n derived from the KC extraction models listed in Table 5 for the SSM_12 dataset, which has 58 KCs. Fig. 6(b) shows the results of applying the same methods to the SSM_13 dataset, which has 24 KCs. Fig. 6 shows that both the OKCE relations and the best set follow the performance pattern of the KC set extracted by Elastic Net (LASSO) + RF in the first step because the feature domain is reduced and then utilized by the Elastic Net (LASSO) + RF.

As shown in Table 5 and Fig. 6(a), the OKCE model can extract KC subsets that have an impact on predicting the learning state of KCs from a given feature domain, thereby providing the best approach for assessing the performance of the KC subset. Fig. 6(b) shows that the DKT performance of the OKCE relations is generally better than that of the entire KC set. This could be attributed to the fact that KC subset size was fixed. That is, because the number of KCs in SSM_13 is 24, and the number of KCs in SSM_12 is 58, which is twice as large as SSM_13, five KC pairs might be insufficient to determine the best performance. Therefore, if the size of the KC subset to be extracted by the OKCE model is adjusted

TABLE 3. Titles for KCs in SSM12.

KC ID	KC Title
KC01	(proper fraction) + (proper fraction) with different denominators without regrouping
KC02	(proper fraction) + (proper fraction) with different denominators with regrouping
KC03	(proper fraction) + (mixed fraction), (mixed fraction) + (proper fraction) and (mixed fraction) + (mixed fraction) with different denominators
KC04	(proper fraction) + (proper fraction) with different denominators without regrouping - intro
KC05	(proper fraction) + (proper fraction) with different denominators with regrouping
KC07	Addition of various fractions
KC08	(proper fraction) − (proper fraction) with different denominators
KC16	(improper fraction) × (natural number), (mixed fraction) × (natural number)
KC21	(improper fraction) × (proper fraction), (mixed fraction) × (proper fraction)
KC23	Multiplication of various fractions
KC24	Multiplication of three fractions
KC26	(proper fraction) ÷ (proper fraction)
KC27	(improper fraction) ÷ (natural number), (mixed fraction) ÷ (natural number)
KC28	(natural number) ÷ (proper fraction)
KC29	(natural number) ÷ (improper fraction), (natural number) ÷ (mixed fraction)
KC35	Division of three fractions

TABLE 4. DKT performance difference by number of KCs.

Number of KCs	Average of overall AUCs	Average of best AUC by KC	Average of worst AUC by KC	best AUC	worst AUC
$n = 5$	0.756	0.809	0.713	**0.874**	0.654
$n = 10$	0.794	0.810	0.775	0.848	0.750

TABLE 5. DKT performance comparison of feature selection algorithm.

Dataset	Average AUC using DKT trained with			
	Top 5 KCs selected by Elastic Net (LASSO) + RF only	Optimal subset of 5 KCs by OKCE Model	Best subset from total combination of 5 KCs	Entire set of KCs
SSM_11	0.733	0.800	0.806	0.830
SSM_12	0.822	0.845	0.846	0.857
SSM_13	0.798	0.862	0.872	0.829
KDD_RATIO	0.711	0.829	0.840	0.782

flexibly according to the KC size of the future dataset, better performance is expected.

C. PERFORMANCE WITH OKCE MODEL

The optimal KC subset for each KC extracted by the OKCE model can derive meaningful KC pairs that improve

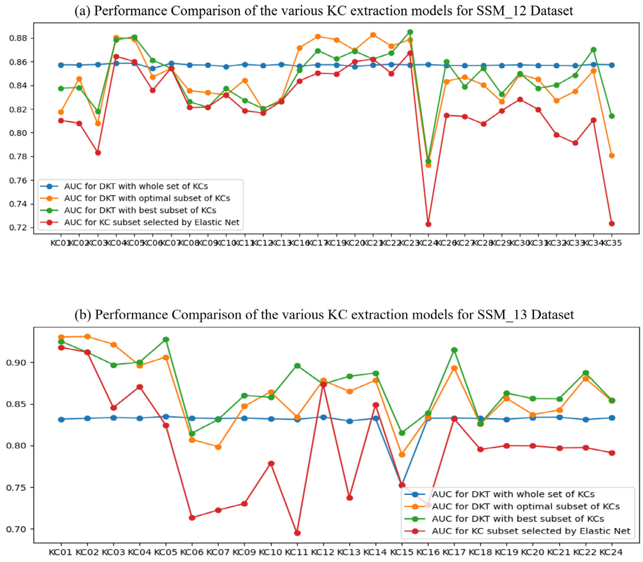


FIGURE 6. AUC Performance of the four DKT models for each KC in the SSM_12 and SSM_13 datasets. KC relations (red line) reduced by Elastic Net (LASSO) + RF (first step of the OKCE model) already show lower performance than AUC for the whole set (blue line). The optimal KC subset obtained in the OKCE model shows a similar pattern to the Elastic Net (LASSO) + RF method, but the actual AUC value is much improved (orange line) and matches the AUC performance for the best set (Green Line).

knowledge tracing performance. If these are OKCE relations, they can be defined as shown in (5):

$$OKCE\ relations = \{(KC_j\ KC_i)|\ KC_i \in \text{subset for } KC_i \text{ by OKCE model for each } KC_i \text{ in Dataset}\} \tag{7}$$

In SSM11-13, KC maps (sets of significant KC relations) predefined by the commercial service provider were included. To verify the performance of the OKCE model, the f1-score of the OKCE relations was calculated for the target KC relations provided by the dataset. However, the f1-score for KDD_RATIO was excluded because a predefined KC map was not provided. As shown in Table 6, for all datasets the f1 score of the OKCE relations is higher than the f1 score of the significant KC pairs obtained in the first step of the OKCE model (Elastic Net (LASSO) + RF). In particular, the f1 score of the OKCE model was higher than that of the top five relations in KC pairs derived by Elastic Net (LASSO) + RF. Therefore, the two stepwise procedures of the OKCE model with the dual-net based on DKT improve the performance of significant KC pair extraction from the learning data compared to the existing machine learning method.

To further confirm this, Fig. 7 shows a heatmap of the OKCE relations, Elastic Net (LASSO) + RF with five KC pairs, and Elastic Net (LASSO) + RF with 10 KC pairs for the KC pairs extracted by the OKCE model at each step. Heatmaps of the target KC relations for SSM_11-13 were also presented for comparison.

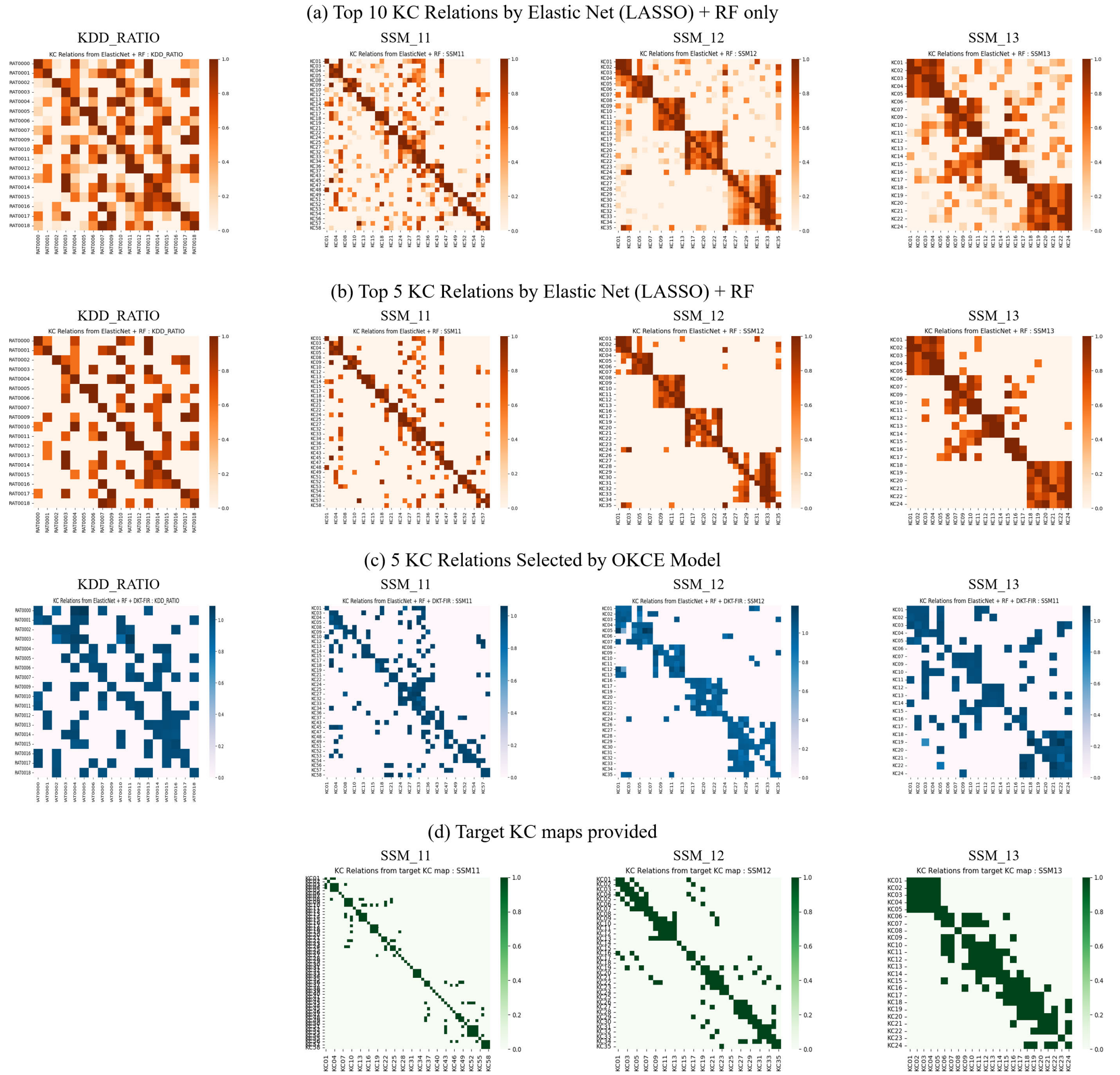


FIGURE 7. Heatmap for selected KC relations for all datasets by (a) Elastic Net (LASSO) + RF only (b) Elastic Net (LASSO) + RF (top 5 relations) (c) OKCE model (significant 5 relations) (d) target KC maps provided by content service provider.

TABLE 6. ***f1 score*** **of OKCE relations to target KC relation.**

Dataset	*f1 score*		
	Elastic Net (LASSO) + RF	Top 5 relations of Elastic Net (LASSO) + RF	OKCE Model
SSM_11	0.1663	0.2656	**0.33**
SSM_12	0.2606	0.3459	**0.4713**
SSM_13	0.3188	0.4458	**0.4828**

VI. CONCLUSION

The OKCE model proposed in this paper extracts the optimal neighbor KC set highly correlated to a KC using an Elastic Net (LASSO) + RF and Dual-net method, so that the knowledge-concept graph can be completed according to the learning state of the learner.

The advantage of the OKCE model is that it narrows the scope of the KC domain by selecting the correlated neighboring KCs for each KC using multiple regression analysis. In addition, the optimal KC subset that predicts the learning state of a specific KC with a high probability can be extracted through the DKT-based dual-net. With the optimal KC subset obtained in this manner, the KC map that reflects the correlation between concepts can be presented more accurately and thus used for personalized learning. For example, a learner who has acquired a specific concept can be guided to the next correlated concept as recommended by the OKCE model. The OKCE model can accommodate both the explicability of machine learning and the accuracy of deep learning, as ML methods such as Elastic Net (LASSO) + RF would reduce the features, and a deep learning-based dual-net would select the features. In addition, the generation and processing times of the model can be reduced because the data size used for the DKT method is reduced.

However, the size n of the final KC set was dedicated to five and ten KCs, which were arbitrarily selected during the pre-processing of the KDD Cup data, which is the subject of the experiment. In this study, the size n of the final KC subset was set to five in the pre-processing process for the dataset of KDD Up, the test subject. However, if the number of optimal neighbor KCs is determined in proportion to the size of the original dataset and the number of KCs, the optimized KC relation set can be extracted to support a more accurate and efficient learning prediction.

The OKCE model suggested in this study can support completion of the knowledge map by deriving the optimal relationships among knowledge concepts, and is expected to provide a more adaptive learning path for each student. In order to enhance the research, verifying the validity of the OKCE model in practice is necessary. For this, additional research on analyzing whether the learning effect increases when actually applying this model to a commercial service should be conducted. In addition, the OKCE model can be extended to perform as a useful learning analysis tool. Applying the OKCE model, the research on the effect of personalized learning between in the online education environment enforced by COVID-19 and in the face-to-face education environment after COVID-19 can be considered. This would be expected to show a comparison of the students' educational levels before and after the pandemic circumstances.

Therefore, this study considers the main goals of future research as follows:

- Improving OKCE model: comparison and analysis of the cases where the number of neighboring KCs for each KC and the threshold value are proportionally selected
- Applying the OKCE model in practice: analysis of the learning effect for learners when applied to the commercial educational services
- Analyzing with the OKCE model: comparison of the learning effect for learners when the OKCE model is applied to non-face-to-face and face-to-face learning environments

REFERENCES

[1] H. G. Oh, "Analysis of major social changes and information security issues after COVID-19," *Commun. Korean Inst. Inf. Scientists Eng.*, vol. 38, no. 9, pp. 48–56, 2020. [Online]. Available: https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09910486

[2] K. Lee, S. Kwon, C. Yang, D. Koh, K. Kim, and M. Choi, "Exploring future school education scenarios in the era of digital transformation," *Korean J. Teacher Educ.*, vol. 37, no. 2, pp. 1–25, 2021. [Online]. Available: http://www.earticle.net/Article.aspx?sn=393898

[3] M. Zhang, J. Zhu, Z. Wang, and Y. Chen, "Providing personalized learning guidance in MOOCs by multi-source data analysis," *World Wide Web*, vol. 22, no. 3, pp. 1189–1219, May 2019, doi: 10.1007/s11280-018-0559-0.

[4] A. Shemshack and J. M. Spector, "A systematic literature review of personalized learning terms," *Smart Learn. Environ.*, vol. 7, no. 1, pp. 1–20, Dec. 2020, doi: 10.1186/s40561-020-00140-9.

[5] D. Lee, Y. Huh, C.-Y. Lin, and C. M. Reigeluth, "Technology functions for personalized learning in learner-centered schools," *Educ. Technol. Res. Develop.*, vol. 66, no. 5, pp. 1269–1302, Oct. 2018doi: 10.1007/s11423-018-9615-9.

[6] M. Brown, M. McCormack, J. Reeves, D. C. Brook, D. S. Grajek, B. Alexander, and N. Weber, "2020 educause horizon report teaching and learning edition," Educause, Tech. Rep., 2020, pp. 14–16. [Online]. Available: https://library.educause.edu/-/media/files/library/2020/3/2020_horizon_report_pdf.pdf?la=en&hash=08A92C17998E8113BCB15DCA7BA1F467F303BA80

[7] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Model. User-Adapted Interact.*, vol. 4, no. 4, pp. 253–278, 1995, doi: 10.1007/bf01099821.

[8] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Cambridge, MA, USA, 2015, pp. 505–513. [Online]. Available: https://dl.acm.org/doi/10.5555/2969239.2969296

[9] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian, "Prerequisite-driven deep knowledge tracing," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Singapore, Nov. 2018, pp. 39–48, doi: 10.1109/ICDM.2018.00019.

[10] R. S. J. D. Baker, A. T. Corbett, and V. Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing," in *Proc. Int. Conf. Intell. Tutoring Syst.*, 2008, pp. 406–415, doi: 10.1007/978-3-540-69132-7_44.

[11] Z. A. Pardos and N. T. Heffernan, "KT-IDEM: Introducing item difficulty to the knowledge tracing model," in *Proc. Int. Conf. User Model., Adaptation, Personalization*, Girona, Spain, 2011, pp. 243–254, doi: 10.1007/978-3-642-22362-4_21.

[12] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, vol. 26. New York, NY, USA: Springer, 2013, p. 13. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4614-6849-3

[13] I. O. Bohachevsky, M. E. Johnson, and M. L. Stein, "Generalized simulated annealing for function optimization," *Technometrics*, vol. 28, no. 3, pp. 209–217, 1986. [Online]. Available: https://doi.org/10.2307/1269076

[14] T. Weise. (2009). *Global Optimization Algorithms: Theory and Application*. Self-Published Thomas Weise. [Online]. Available: http://www.it-weise.de/projects/book.pdf

[15] C. Ambroise and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 10, pp. 6562–6566, 2002. [Online]. Available: https://www.jstor.org/stable/3058705

[16] H. Choi and M. Lee, "Analysis of prerequisite relation in knowledge graph using ARM and MSMM: Focusing on problem evaluation data of K-12 math," *J. Digit. Contents Soc.*, vol. 23, no. 6, pp. 1131–1140, Jun. 2022, doi: 10.9728/dcs.2022.23.6.1131.

[17] N. T. Son, J. Jaafar, I. A. Aziz, and B. N. Anh, "Meta-heuristic algorithms for learning path recommender at MOOC," *IEEE Access*, vol. 9, pp. 59093–59107, 2021, doi: 10.1109/ACCESS.2021.3072222.

[18] Y. Lu, P. Chen, Y. Pian, and V. W. Zheng, "CMKT: Concept map driven knowledge tracing," *IEEE Trans. Learn. Technol.*, vol. 15, no. 4, pp. 467–480, Aug. 2022, doi: 10.1109/TLT.2022.3196355.

[19] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.

[20] K.-W. Chang, W.-T. Yih, B. Yang, and C. Meek, "Typed tensor decomposition of knowledge bases for relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1568–1579, doi: 10.3115/v1/d14-1165.

[21] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2D knowledge graph embeddings," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2018, pp. 1–8, doi: 10.1609/aaai.v32i1.11573.

[22] I. Balazevic, C. Allen, and T. M. Hospedales, "Hypernetwork knowledge graph embeddings," in *Proc. Int. Conf. Artif. Neural Netw.*, Munich, Germany, 2019, pp. 553–565, doi: https://doi.org/10.1007/978-3-030-30493-5_52.

[23] J. Hipp, U. Guntzer, and G. Nakhaeizadeh, "Algorithms for association rule mining—A general survey and comparison," *ACM SIGKDD Explor. Newslett.*, vol. 2, no. 1, pp. 58–64, Jun. 2000.

[24] R. Shi and S. A. Conrad, "Correlation and regression analysis," *Ann. Allergy, Asthma Immunology*, vol. 103, no. 4, pp. 35–41, 2009. [Online]. Available: http://college.acaai.org/sites/default/files/Statistics%20for%20Clinicians%20-%20Annals%20Suppl%20Oct%2009%20-%20ALL.pdf

[25] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc., B*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[26] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, Feb. 1966, doi: 10.1126/science.153.3731.34.

[27] T. Van Der Ploeg and E. W. Steyerberg, "Feature selection and validated predictive performance in the domain of *Legionella pneumophila*: A comparative study," *BMC Res. Notes*, vol. 9, no. 1, pp. 1–7, Dec. 2016.

[28] D. V. Hinkley and S. Shi, "Importance sampling and the nested bootstrap," *Biometrika*, vol. 76, no. 3, pp. 435–446, 1989.

[29] R. J. Lewis, "An introduction to classification and regression tree (CART) analysis," in *Proc. Annu. Meeting Soc. Academic Emergency Med.*, San Francisco, CA, USA, 2000, p. 14. [Online]. Available: https://www.researchgate.net/profile/Roger-Lewis-8/publication/240719582_An_Introduction_to_Classification_and_Regression_Tree_CART_Analysis/links/0046352d3fb18f1740000000/An-Introduction-to-Classification-and-Regression-Tree-CART-Analysis.pdf

[30] Y. Liu, Y. Wang, and J. Zhang, "New machine learning algorithm: Random Forest," in *Proc. Int. Conf. Inf. Comput. Appl.*, Chengde, China, 2012, pp. 246–252, doi: 10.1007/978-3-642-34062-8_32.

[31] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.

[32] H. Zou, "The adaptive lasso and its Oracle properties," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006, doi: 10.1198/016214506000000735.

[33] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005, doi: 10.1109/tkde.2005.50.

[34] F. Lu and E. Petkova, "A comparative study of variable selection methods in the context of developing psychiatric screening instruments," *Statist. Med.*, vol. 33, no. 3, pp. 401–421, Feb. 2014, doi: 10.1002/sim.5937.

[35] M. Wojtas and K. Chen, "Feature importance ranking for deep learning," 2020, *arXiv:2010.08973*.

[36] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003. [Online]. Available: https://dl.acm.org/doi/10.5555/944919.944968

[37] S. Yang, P. Wu, and H. Guo, "DualNet: Locate then detect effective payload with deep attention network," in *Proc. IEEE Conf. Dependable Secure Comput. (DSC)*, Aizuwakamatsu, Japan, 2021, pp. 1–8.

[38] A. P. S. S. Raj and S. K. Vajravelu, "DDLA: Dual deep learning architecture for classification of plant species," *IET Image Process.*, vol. 13, no. 12, pp. 2176–2182, Sep. 2019, doi: 10.1049/iet-ipr.2019.0346.

[39] T. Kaser, S. Klingler, A. G. Schwing, and M. Gross, "Dynamic Bayesian networks for student modeling," *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 450–462, Oct./Dec. 2017, doi: 10.1109/TLT.2017.2689017.

[40] P. I. Pavlik Jr., H. Cen, and K. R. Koedinger, "Performance factors analysis—A new alternative to knowledge tracing," in *Proc. Conf. Artif. Intell. Educ., Building Learn. Syst. Care, Knowl. Represent. Affect. Model.*, Brighton, U.K., 2009, pp. 531–538. [Online]. Available: https://eric.ed.gov/?id=ED506305

[41] R. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Comput.*, vol. 2, no. 4, pp. 490–501, 1990, doi: 10.1162/neco.1990.2.4.490.

[42] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306, doi: 10.1016/j.physd.2019.132306.

[43] C. B. Sahin and B. Diri, "Robust feature selection with LSTM recurrent neural networks for artificial immune recognition system," *IEEE Access*, vol. 7, pp. 24165–24178, 2019, doi: 10.1109/ACCESS.2019.2900118.

[44] D. Kaur, R. Kumar, N. Kumar, and M. Guizani, "Smart grid energy management using RNN-LSTM: A deep learning-based approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[45] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110120, doi: 10.1016/j.chaos.2020.110120.

[46] O. R. Bininda-Emonds, K. E. Jones, S. A. Price, M. Cardillo, R. Grenyer, and A. Purvis, "Garbage in, garbage out," *Phylogenetic Supertrees*, vol. 4, pp. 267–280, Jan. 2004.

[47] J. Schmid, "The relationship between the coefficient of correlation and the angle included between regression lines," *J. Educ. Res.*, vol. 41, no. 4, pp. 311–313, Dec. 1947, doi: 10.1080/00220671.1947.10881608.

[48] D. E. Farrar and R. R. Glauber, "Multicollinearity in regression analysis: The problem revisited," *Rev. Econ. Statist.*, vol. 49, no. 1, pp. 92–107, 1967. [Online]. Available: https://doi.org/10.2307/1937887

[49] D. Debeer and C. Strobl, "Conditional permutation importance revisited," *BMC Bioinf.*, vol. 21, no. 1, pp. 1–30, Dec. 2020.

[50] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinform.*, vol. 9, no. 307, pp. 1–11, 2008, doi: 10.1186/1471-2105-9-307.

[51] K. N. Carter, K. Kruse, T. Blakely, and S. Collings, "The association of food security with psychological distress in New Zealand and any gender differences," *Social Sci. Med.*, vol. 72, no. 9, pp. 1463–1471, May 2011, doi: 10.1016/j.socscimed.2011.03.009.

[52] K. Kiernan and K. E. Pickett, "Marital status disparities in maternal smoking during pregnancy, breastfeeding and maternal depression," *Social Sci. Med.*, vol. 63, no. 2, pp. 335–346, Jul. 2006, doi: 10.1016/j.socscimed.2006.01.006.

[53] M. Afeef, "Analyzing the impact of working capital management on the profitability of SME's in Pakistan," *Int. J. Bus. Social Sci.*, vol. 2, no. 1, pp. 173–183, 2011. [Online]. Available: https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=2890773

HYUNHEE CHOI received the M.S. degree in statistics from Korea University, Seoul, South Korea, in 1997, and the Ph.D. degree in statistics from Hoseo University, Chungnam, South Korea, in 2019.

Since 1996, she has been a Researcher of applied statistics. She is currently with LAIVDATA—AI-based Edtech startup company, South Korea. She is the author of three books and several other articles. Her research interests include data analysis, learning path generation, data science, adaptive learning, and edtech.

HAYUN LEE was born in Seoul, South Korea, in 2000. She is currently pursuing the B.S. degree with the School of Computer Science and Engineering, Chung-Ang University, Seoul.

She has been working as a Researcher with the Edutech Research and Development Center, LAIVDATA—AI-based edtech startup company, South Korea. Her research interests include CSE, machine learning, pattern recognition, and AI-based educational technology.

MINJEONG LEE was born in South Korea. She received the B.S. degree in computer science and engineering from Chung-Ang University, in 1994, and the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1996. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Korea University, Seoul, South Korea.

She has been a Software Engineer and the Project Manager of the LG's Research Center and for C&S Technology for 15 years, since 1996, and she also has been with the Developer Relations Team, Samsung Software Center, South Korea, for five years, since 2000. She has been with the Da Vinci College of General Education, Chung-Ang University, since 2016. She has been the Director of the Edutech Research and Development Center, LAIVDATA, since 2021. Her research interests include computing education, digital literacy, computational thinking, intelligent tutoring systems, and AI educational technology.