# Multi-relational Matrix Factorization Approach for Educational Items Clustering

Denon Arthur Richmond Gono*, Bi Tra Goore, Yves Tiecoura, and Kouamé Abel Assielou

***Abstract*—Pedagogical models development requires several steps, one of which is the mapping of tasks and skills, also known as the educational items clustering. This activity of clustering educational items usually requires the participation of domain experts. However, discovering the exact skills involved in performing the tasks is a complex activity for them. This paper aims at solving the task and skill-mapping problem by proposing an approach based on the Weighted Multi-Relational Matrix Factoring technique to help experts in this task. This approach relies on two types of relationship, the " student does task" relationship and the "student has skills" relationship through a latent factor model to reconstruct the " task requires skill" relationship, the latter being the mapping between tasks and skills. An evaluation conducted on a group of two hundred (200) students in lower 6th class in a general secondary school (Côte d'Ivoire), showed that this approach brought an improvement rate of about 82.8% of the skill-task mapping proposed by the experts in the field. This result confirms that our approach not only allows us to map tasks and skills but also to significantly improve the updating of curricula.**

***Index Terms*—Pedagogical models, skills discovery, matrix factorization, Q-Matrix, WMRMF.**

## I. Introduction

With an insertion rate of 14.6% [1] of its graduates, Technical Education and Vocational Training (TEVT) clearly demonstrates the difficulties encountered in the Education - Training sector in Côte d'Ivoire. To address this situation, numerous reforms and studies have been undertaken [1], particularly the switch to the Competency-Based Approach (CBA) as the new educational technology to replace the pedagogy by objectives (PBO).

The introduction of the CBA in General and Technical Education and Vocational Training in Côte d'Ivoire to address the mismatch between employment and training. To this end, one of the foundations of the CBA is the learning context, which must be as close as possible to that of the professional environment in which the future graduate will have to work. Also, the changeover to this new pedagogical approach makes it possible to improve curricula [2].

The discovery of the competencies behind the tasks is generally an activity dedicated to the domain experts. In the Competency-Based Approach (CBA), this activity of mapping competency elements to tasks, which results in a training repository or learning model, very often requires the participation of several domain experts.

Moreover, the pedagogical scenarization, which is one of the important tasks in learning environments, also requires this mapping of skills and tasks. However, the mapping of competency skills and the resulting tasks remains a complex activity given the large amount of educational data in traditional systems such as learning environments. According to Barnes [3] and Cen and Koedinger *et al*. [4] mapping items to latent skills is a notoriously difficult task. Thus, the proposal of new approaches to help experts in this task is a major line of study.

This paper aims to propose a new approach to clustering educational data based on the Weighted Multi-Relational Matrix Factoring technique. Unlike the work done in the state of the art, this approach exploits not only the different domain relations but also the weight factor of each of its relations to predict the mapping between skills and tasks.

## II. State of the Art

The mapping of tasks and skills in an education system aims to assess the mastery of skills by learners. This mapping of tasks and skills, also called Q-Matrix, is a skills repository that takes the form of a job-skill or task-skill mapping requires a skilled workforce and significant expertise [5], [6].

In literature, there are two approaches to mapping tasks and skills: model-based approaches and similarity measure-based approaches [7].

Approaches using similarity measures are based on the assumption that students will tend to perform similarly on items requiring the same skill; thus seeking to identify the similarity between pairs of items. These approaches first calculate the similarity for each pair of items. Then, the result obtained can be used to group the items. Choi *et al*., [8] have identified in a study, a hundred measures of similarity by domain of competence. This study shows a strong correlation between the existing measures. Also, for them, the accuracy of the mapping is related to the choice of the similarity measure adopted. Following this work, authors such as J. Řihák and R. Pelánek [9], identified seven (7) measures of similarity implemented in the educational context. The results of their study showed that Cohen's Kappa method gave better results in terms of skill discovery. In this context, Nazaretsky and Hershkovitz *et al*. [10] proposed a new similarity measure called Kappa Learning to improve the Kappa method. Unlike Kappa, Kappa Learning takes into account the notion of learning.

As for the model-based approaches, they reduce the dimensionality of the problem and try to deduce the latent factors underlying the tasks. S. Lakshmi Prabha *et al*., [11] propose a model of Learning Factor Analysis, a model combining statistics, human expertise and combinatorial research to evaluate and improve a cognitive model but above all to explore educational data.

Manuscript received June 19, 2022; revised August 13, 2022; accepted August 31, 2022.

The authors are with Research Laboratory in Computer Science and Telecommunications (LARIT) at the Institut National Polytechnique Felix HOUPHOUET Boigny (INP-HB) Yamoussoukro, Côte d'Ivoire.

*Correspondence: arthurgono@gmail.com

doi: 10.18178/ijiet.2023.13.1.1778

Recently N. Thai-Nghe *et al.*, [12] to predict student's performance in Intelligent Tutorial Systems, have used a technique from recommendation systems, namely Matrix Factorization. Authors such as M. C. Desmarais *et al.*, [13] for the clustering of educational items have exploited a tendency of this approach called ALS (Alternative Least Squares Factorization). This approach exploits the "student performs task" relationship.

However, approaches based on similarity measures and those based on models have a major limitation. Indeed, they only exploit the stduent's performance in order to map tasks and skills.

Nevertheless, the Matrix Factorization technique has encountered several improvements such as: MRMF (Multi-Relational Matrix Factorization) [14], WMRMF (Weighted Multi-Relational Matrix Factorization) [15], So-WMRMF (Social Weigthed Multi-Relational Matrix Factorization) [16], Emo - WMRMF (Emotional Weigthed Multi-Relational Matrix Factorization) [15], SoEmo-WMRMF (Socio-Emotional Weigthed Multi-Relational Matrix Factorization) [17]. These approaches generally try to draw on several domain relationships. However, they have been used just for predicting student performance in Intelligent Tutorial Systems.

This study aims to solve the problem of mapping tasks and skills. To achieve this, we propose an approach based on the Weighted Multi-Relational Matrix Factorization technique to help experts in this task. This approach relies on two types of relationship, namely the "learner performs task" relationship and the "learner has acquired" relationship through a latent factor model to reconstruct the "task requires skill" relationship, this last relationship being the mapping between tasks and skills.

## III. Related Work

The present study aims to explore the Multi-Relational Matrix Factorization technique in the field of skills and tasks (items) mapping. Therefore, this session describes the working principle of the Multi-Relational Matrix Factoring technique.

In the classical FM approach, we considers $S$ a set of student, $I$ a set of tasks and $P$ a range of possible performance scores [18]. The principle of this technique is to find two small matrices $W_1$ ( students) and $W_2$ (tasks) such that the different performances (Matrix R) achieved by the students can be approximated by relation (1) [19]

$$R \approx W_1 \; W_2^T \tag{1}$$

In equation (1), $W_1 \in \mathbb{R}^{S \times F}$ describe a matrix in which each row *s* represents a vector that contains $F$ latent factors best describing the profile of $S$ student and $W_2 \in \mathbb{R}^{I \times F}$ a matrix where each row $i$ is a vector containing $F$ latent factors describing the task $i$ . [16]. The prediction of the performance of $S$ students for $i$ task to be performed is given by the equation (2):

$$\hat{p}_{si} = \sum_{f=1}^{F} w_{1_{sf}} w_{2_{if}} = w_{1_s} w_{2_i}^T \tag{2}$$

In this technique, we consider $N$ types of entities $\{E_1,...,E_N\}$ connected by $M$ types of relationships $\{R_1,...,R_M\}$ that can be strongly correlated to each other [20]. The matrices $W_1, W_2, ..., W_n$ ($n \in N$), designate the parameters of the model. These parameters are generally learned by optimizing the objective function (3) by the technique of stochastic gradient descent [21].

$$o^{WMRMF} = \sum_{r=1}^{M} \Theta_r \sum_{(s,i)\in R_r} \left( R_{r_{si}} - w_{r1_s} w_{r2_i}^T \right)^2 + \lambda \left( \sum_{n=1}^{N} \| W_n \|_F^2 \right) \tag{3}$$

with $R_r = \{(E_{1r}; E_{2r})\} (r = 1...M)$

The prediction error $e_{si}$ (see equation 4) is calculated by taking the difference between the actual $R_{si}$ performance value and the predicted performance $\hat{p}_{si}$ value for each pair (*s*, *i*).

$$e_{si} = \left( R_{si} - w_{1_s} w_{2_i}^T \right) \tag{4}$$

In Eq. (3), $\| \;.\; \|_F^2$ denotes the Frobenius norm and $\lambda$ is a regularization parameter [19]. Equation (5) gives the function representing the weight factor for the different relations of the domain:

$$\Theta_r = \begin{cases} 1, & \text{if } r \textit{ is the main relationship} \\ \theta, & si \; (0 < \theta < 1) \end{cases} \tag{5}$$

The parameters of WMRMF model are updated through Eqs. (6) and (7) [22]:

$$w'_{r1_{sk}} = w_{r1_{sk}} + \beta \left( 2\Theta_r e_{r_{si}} w_{r2_{ik}} - \lambda w_{r1_{sk}} \right) \tag{6}$$

$$w'_{r2_{ik}} = w_{r2_{ik}} + \beta \left( 2\Theta_r e_{r_{si}} w_{r1_{sk}} - \lambda w_{r2_{ik}} \right) \tag{7}$$

In these equations, the factor $\beta$ designates the learning rate.

## IV. Proposed Approach

### A. Problem Formulation

Our approach to mapping skills and items is based on a Weighted Multi-Relational Matrix Factorization model as described in Section III. The objective of this study being to predict the Q-Matrix of the experts, we also consider the following three relations: the relation 1) "*learner performs task*", the relation 2) "*learner to achievements*" and the relation 3) "*task requires skill*".

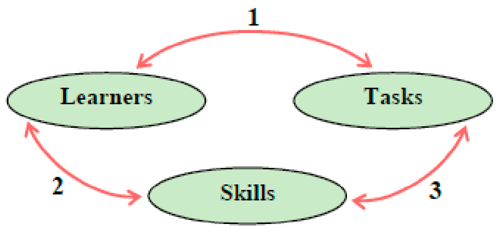



Fig. 1. Relationships taken into account in the proposed approach.

The relation (1) is represented in the form of a matrix of results $R \in \mathbb{R}^{S\times I}$. In this matrix, each pair (*s*,*i*) designates the score obtained by $S$ student when performing $I$ task. To perform $I$ task, the student must use the different skills required to perform this task, implemented in relation (2).

The relation (2) is represented as a matrix $A \in {}^{S\times K}$ of learning outcomes. K is the set of skills. Matrix A is a mapping between learners and the different skills to be acquired. It defines the level of mastery of the different competences implemented by the experts of the underlying educational domain.

The relation (3) is represented by a matrix $Q \in \square^{K\times I}$, similar to the Q-Matrix as proposed by domain experts. It establishes a map between the tasks and the different skills useful for the execution of these tasks. When $Q_{ij} = 1$, it means that execution of task *j* requires skill *i*. if $Q_{ij} = 0$, then execution of task *j* does not require skill *i*. In the present study, the execution of a task may require one or several skills.

The principle of our approach is to predict the Q- Matrix of the experts and compare the similarity between them. This similarity will allow us to adjust or refine the Q-Matrix of the experts.

### *B. Q-Matrix Elaboration by the Expert*

To predict the Q-Matrix of Experts, our approach uses the WMRMF factorization technique based on the $R$ and $A$ matrices. The principle is to perform supervised learning based on the $R$ and $A$ matrices data in order to predict the Expert Q-Matrix using equation (2). The system takes advantage of the interaction of the two relations "*learner performs task*" and "*learners has acquired*". In this paper, we conducted a survey of learners in the second cycle of general education, enrolled in the Première class, at Lycée Moderne Khalil (Daloa, Côte d'Ivoire) during the 2021-2022 school year.

This study is based on the educational program and implementation guide provided by the Ministry of National Education and Literacy of Côte d'Ivoire to all teachers in the learning context. This program and guide defines, for all levels of education, the output profile, the disciplinary field, the pedagogical regime and it presents the body of the discipline program.

The body of the program is broken down into several components: skill, theme, lesson, an example of a course application situation, and the instructional content. In this study, we used the following lessons: limits and continuity, extension of the notion of limit, and derivation. An extract of the skills related to these lessons is presented in Table I. This study includes 84 skills.

TABLE I: EXTRACT FROM SOME SKILLS USED IN THIS STUDY

| Skill 1: deal with a situation related to algebraic calculations and functions | | | |
|---|---|---|---|
| Lesson: Derivation | | | |
| *C* | *Kc* | Skills | Contents |
| 1 | 001 | Knowing | The definition of the drift number at a point of a function |
| 1 | 002 | Knowing | The definition of the derivative function on an open interval. |
| 1 | 003 | Note | The derivative of a function on an open interval |
| 1 | 004 | Determine | An extremum of a function using its derivative. |
| 1 | 005 | Calculate | the derivative number of a function at a point |
| … | … | … | … |
| 1 | 021 | Dealing with a situation | Using derivation |
| … | … | … | … |

To develop the Q-matrix for evaluating the weighted multi-rational matrix factorization approach, we appeal to teachers (11). For each skill, they proposed activities (exercises) to evaluate the learners. An example of activities is given in Table II where for the knowledge components $Kc_1$ "*Knowing the number of p-tuples of a set with n elements*" the activities Q1 and Q2 are proposed. In total, these experts proposed 200 activities for three (3) lessons.

TABLE II: EXTRACT FROM SOME TASKS PROPOSED BY TEACHERS

<table>
<tr><th colspan="8">ACTIVITIES</th></tr>
<tr><td colspan="8">Knowing the number of p-tuples of a set with n items</td></tr>
<tr><td colspan="4">Q1: The number of tuples of a set with n elements is :</td><td colspan="4">Q2: Let the set $E = \{1;2\}$ . The number of triples of the set E is :</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>a</td><td>b</td><td>c</td><td>d</td></tr>
<tr><td>$n\times p$</td><td>$n+p$</td><td>$n^p$</td><td>$p^n$</td><td>8</td><td>6</td><td>3</td><td>0</td></tr>
<tr><td colspan="8">Know the number of p-element arrangements of an n-items set (n ≥ p)</td></tr>
<tr><td colspan="4">Q3: The number of p-arrangements of a set with n elements is :</td><td colspan="4">Q4: Let the set $E = \{a;b;c;d;e\}$ . The number of arrangements of 2 elements E is :</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>a</td><td>b</td><td>c</td><td>d</td></tr>
<tr><td>$A_n^n$</td><td>$A_n^p$</td><td>$n^p$</td><td>$A_p^p$</td><td>24</td><td>16</td><td>8</td><td>20</td></tr>
</table>

These teachers were also committed, through several working sessions, to propose a first version of the mapping between the activities and the skills provided by the supervisory ministry (Ministry of Education and Literacy).

This first version of the Q-matrix was then validated on the basis of the principle of prioritization of tasks and skills as described in the work of authors Villanueva *et al*. and Kikumi K. Tatsuoka [23], [24]. This validation is done by proposing a Q-Matrix tree diagram in order to verify if each skill fits into the hierarchy of useful concepts to be taught to student.

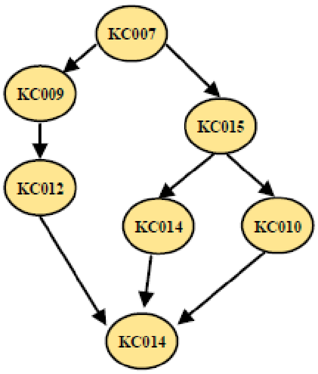



Fig. 2. Competency hierarchy diagram.

The main purpose of this study was to obtain a performance matrix and a learning matrix and to develop the Q-Matrix of the experts, which will allow us to calculate the similarity with the predicted Q-Matrix. Based on this study, a Q-Matrix was developed by these teachers, experts of the domain.

Tasks

| Skills | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_5$ | $Q_6$ | $Q_7$ |
|---|---|---|---|---|---|---|---|
| $KC_1$ | | 1 | | 1 | | | |
| $KC_2$ | | | 1 | | | 1 | 1 |
| $KC_3$ | 1 | | 1 | | | | |
| $KC_4$ | | | | 1 | 1 | | |
| $KC_5$ | 1 | | 1 | | | | |
| $KC_6$ | | 1 | | | | 1 | 1 |

Fig. 3. Extract from the Q-Matrix of experts for 6 skills and 7 tasks.

This study also grouped 200 learners to evaluate 200 tasks in the mathematics discipline: 40,000 performances (Table III).

TABLE III: INFORMATION ON LEARNERS, TASKS AND SKILLS

| # Learners | #Tasks | # Skills | # Performance |
|---|---|---|---|
| 200 | 200 | 84 | 40.000 |

Tasks

| Students | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_5$ | $Q_6$ | $Q_7$ |
|---|---|---|---|---|---|---|---|
| $S_1$ | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| $S_2$ | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| $S_3$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $S_4$ | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| $S_5$ | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

Fig. 4. Extract from Matrix R (Performance Matrix) for 5 student and 7 tasks.

Skills

| Students / Apprenants | $KC_1$ | $KC_2$ | $KC_3$ | $KC_4$ | $KC_5$ | $KC_6$ |
|---|---|---|---|---|---|---|
| $S_1$ | 0.1 | 0.8 | 0.9 | 0.2 | 0.5 | 0.7 |
| $S_2$ | 0.6 | 0.1 | 0.9 | 0.8 | 0.8 | 0.7 |
| $S_3$ | 0.5 | 0.4 | 0.7 | 0.6 | 0.9 | 1 |
| $S_4$ | 0.2 | 0.6 | 0.7 | 0.4 | 0.6 | 1 |
| $S_5$ | 0.1 | 0.2 | 0.6 | 0.6 | 0.4 | 0.5 |

Fig. 5. Extract from Matrix A (Acquired Matrix) for 5 students and 6 skills.

Following the different tasks developed by the experts, students were invited to take part in several course sessions and assessments on a platform designed and implemented for this purpose. Students solve problems in the learning system and each interaction between the learner and the system is recorded as a transaction line.

The data collected through the learning environment led to the generation of matrices *R* and *A*, presented respectively by Fig. 4 and Fig. 5.

### *C. Evaluation of Proposition Approach*

To evaluate our approach in the field of skills and tasks mapping, we used two (2) types of metrics: first the Root Mean Squared Error RMSE and the Mean Absolute Error MAE to calculate the prediction error of the predicted Q-Matrix then the Cohen's Kappa similarity measure to determine the difference between the experts' Q-Matrix and the predicted Q-matrix. This difference will make it possible to identify the different pairs (Skills, tasks) which will require interpretation and analysis by the experts.

The different expressions of RMSE, MAE and Cohen's Kappa measure are given respectively through equations (8), (9) and (10).

$$RMSE = \sqrt{\frac{\sum_{(r,s,i)\in D^{test}} \left(p_{si} - \hat{p}_{si}\right)^2}{\left|D^{test}\right|}} \tag{8}$$

$$MAE = \frac{1}{\left|D^{test}\right|} \sum_{(r,s,i)\in D^{test}} \left|p_{si} - \hat{p}_{si}\right| \tag{9}$$

$$Kappa = \frac{Po - Pe}{1 - Pe} \tag{10}$$

With $Po = \sum_{i=1}^{r} p_{ii}$ and $Pe = \sum_{i=1}^{r} p_{i} + P_{+i}$

The evaluation carried out consists in performing a supervised learning using the data of the two matrices *R* and *A* based on the Stochastic Gradient Descent principle and then predicting, when the model is optimized, the experts' Q-Matrix. For the evaluation of this approach, we used a 64-bit Windows working environment, 16 GB of RAM with an Intel Core i5 processor. Our algorithm was simulated in Python.

## V. RESULTS AND DISCUSSION

This section presents the different results of RMSE, MAE and Cohen's Kappa obtained. It also discusses these results while implementing a validation window for these results by the experts.

WMRMF is a technique that not only takes into account different domain relationships but also adds a weight factor that symbolizes the degree of importance of different relationships in order to map skills and tasks. To evaluate this matrix factorization approach, we essentially rely on the *R* and *A* matrices.

The main relation is "learner performs task" with a weight value to 1 and the secondary relation is "learner has acquired" with a weight value equal to 0.9. After evaluation, the parameters that optimize the model have been recorded in Table IV.

TABLE IV: OPTIMIZATION SETTINGS

| WMRMF model parameter | |
|---|---|
| $K = 2$ | **#iter = 41** ; $\beta = 10^{-2}$ ; $\lambda = 0.015$ ; $T = 0:3:28$ ; $\Theta \in \{1;0.9\}$ |

| | |
|---|---|
| $K=3$ | **#iter = 36** ; $\beta=10^{-2}$ ; $\lambda=0.015$ ; $T=0:3:58$ ; $\Theta\in\{1;0.9\}$ |
| $K=4$ | **#iter = 10** ; $\beta=10^{-2}$ ; $\lambda=0.015$ ; $T=0:4:45$ ; $\Theta\in\{1;0.9\}$ |
| $K=5$ | **#iter = 41** ; $\beta=10^{-2}$ ; $\lambda=0.015$ ; $T=0:5:46$ ; $\Theta\in\{1;0.9\}$ |
| $K=6$ | **#iter = 85** ; $\beta=10^{-2}$ ; $\lambda=0.015$ ; $T=0:6:11$ ; $\Theta\in\{1;0.9\}$ |
| $K=7$ | **#iter = 44** ; $\beta=10^{-2}$ ; $\lambda=0.015$ ; $T=0:7:1$ ; $\Theta\in\{1;0.9\}$ |

Fig. 5 gives the evolution of RMSE and MAE errors as a function of the number of iterations. These different prediction errors are calculated by equations (8) and (9) using the Q-Matrix of the experts and the Q-Matrix predicted by the model.

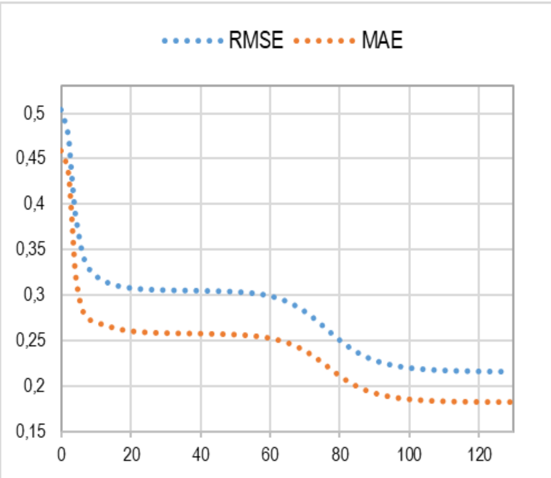


Fig. 6. Evolution of RMSE and MAE by iterations.

We obtain an error RMSE = 0.216 and MAE = 0.182 from 113 iterations. This error represents a form of discrepancy between the Q-Matrix of the experts and the Q-Matrix predicted by the model. This led us to calculate the similarity between these two matrices according to the number of latent factors as presented in Fig. 6.

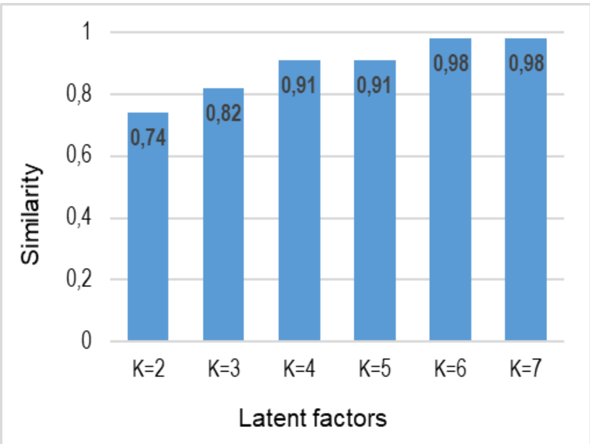
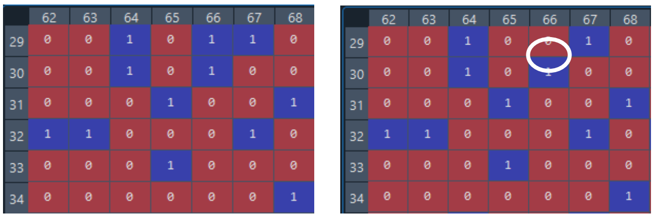


Fig. 7. Similarity between Q-matice and Predicted Q-matrix for different latent factor values.

For $K$= 6 the similarity between the Q-matrix of the experts and the predicted Q-matrix is 0.98, thus denoting 378 irregularities out of 16800 (84×200).

## A. Proposed Approach Validation

Fig. 8a presents the Q-matrix of the experts and Fig. 8b presents the Q-matrix predicted by the WMRMF model. Fig. 8b shows some highlighted irregularities



| | 62 | 63 | 64 | 65 | 66 | 67 | 68 |
|---|---|---|---|---|---|---|---|
| 29 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| 30 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 31 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 32 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 33 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 34 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

a) Q-Matrix of Expert

| | 62 | 63 | 64 | 65 | 66 | 67 | 68 |
|---|---|---|---|---|---|---|---|
| 29 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 30 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 31 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 32 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 33 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 34 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

b) Q-Matrix predicted

Fig. 8. Extract from the Q-Matrix of the experts and the Q-matrix predicted by the model.

The 378 irregularities noted during the evaluation were submitted to the opinion of the experts who designed the Q-matrix. These irregularities are considered propositions of the Q-Matrices based on the matrix learning performed using the *R* and *A* matrices.

It appears from their analysis that, of the 378 proposals made by our matrix factorization approach, these experts consider 313 admissible, i.e. an improvement rate of 82.8%. This result confirms that our approach not only makes it possible to map tasks and skills but also significantly improves the updating of curricula.

## VI. Conclusion

This study proposed an item clustering approach to address the problem of updating training curricula in educational systems. This approach is based on the technique of Weighted Multi-Relational Matrix Factorization. It uses two types of relations, namely the relation “learner performs task” and the relation “learner has knowledge” to construct the relation “task requires competence”. A study carried out on a data set collected in a secondary school showed that, out of 378 proposals made by our approach, these experts deemed 313 acceptable, i.e. an 82.8% improvement rate of the experts' Q-Matrix. This confirms the accuracy of our approach in the field of competence discovery in an educational system.

In future work, integrating external factors such as those related to the classroom, the learner's life, the school, and society could allow us to refine the rate of improvement of the expert Q-Matrix.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Denon Arthur Richmond GONO was the one who carried out the current study and prepared the paper. Bi Tra GOORE was in charge of developing the research methodology and data collection methods. The design phase of the Q-Matrix by domain experts was under the supervision of Kouamé Abel ASSIELOU. The learning environment for data collection was developed under the supervision of Yves TIECOURA, Denon Arthur Richmond GONO who was also the corresponding author. All authors approved the final version.

## References


[1] ETFP, "Réforme de l'Enseignement Technique et la Formation Professionnelle 2016 - 2025," *J. Off. Côte d'Ivoire*, 2016.

[2] Organisation Internationale de la Francophonie, "Les guides méthodologiques d'appui à la mise en œuvre de l'approche par compétences en formation professionnelle," Québec, 2009

[3] B. Tiffany. "The Q-matrix method: Mining student response data for knowledge," American Association for Artificial Intelligence 2005 Educational Data Mining Workshop, AAAI Press, Pittsburgh, PA, USA, 2005.

[4] H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis — A general method for cognitive model evaluation and improvement," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4053, pp. 164–175, 2006, doi: 10.1007/11774303_17.

[5] J. Liu, G. Xu, and Z. Ying, "Data-driven learning of Q-Matrix," *Appl. Psychol. Meas.*, vol. 36, no. 7, pp. 548–564, 2012, doi: 10.1177/0146621612456591.

[6] A. A. Rupp and J. Templin, "The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model," *Educ. Psychol. Meas.*, vol. 68, no. 1, pp. 78–96, 2008, doi: 10.1177/0013164407301545.

[7] R. Pelanek, "Measuring similarity of educational items: An overview," *IEEE Trans. Learn. Technol.*, pp. 1–13, 2019, doi: 10.1109/TLT.2019.2896086.

[8] S. S. Choi, S. H. Cha, and C. C. Tappert, "A survey of binary similarity and distance measures," in *Proc. WMSCI 2009 - 13th World Multi-Conference Syst. Cybern. Informatics, Jointly with 15th Int. Conf. Inf. Syst. Anal. Synth. ISAS 2009 - Proc*., vol. 3, pp. 80–85, 2009.

[9] J. Řihák and R. Pelánek, "Measuring similarity of educational items using data on learners' performance," in *Proc. 10th Int. Conf. Educ. Data Mining, EDM 2017*, pp. 16–23, 2017.

[10] T. Nazaretsky, S. Hershkovitz, and G. Alexandron, "Kappa learning: A New item-similarity method for clustering educational items from response data," in *Proc. 12th Int. Conf. Educ. Data Min*., pp. 129–138, 2019.

[11] S. L. Prabha and A. R. M. Shanavas, "A study on learning factor analysis — An educational data mining technique for student knowledge modeling," *IOSR J. Comput. Eng. Ver. IV*, vol. 17, no. 6, pp. 2278–661, 2015, doi: 10.9790/0661-176495101.

[12] N. Thai-Nghe, L. Drumond, T. Horváth, A. Krohn-Grimberghe, A. Nanopoulos, and L. Schmidt-Thieme, "Factorization techniques for predicting student performance," *Educ. Recomm. Syst. Technol. Pract. Challenges*, pp. 129–153, 2011, doi: 10.4018/978-1-61350-489-5.ch006.

[13] M. C. Desmarais and R. Naceur, "A matrix factorization method for mapping items to skills and for enhancing expert-based," doi: 10.1007/978-3-642-39112-5_45.

[14] K. A. Assielou, C. T. Haba, T. L. Kadjo, K. D. Yao, and B. T. Goore, "Multi relational and social influence model for predicting student performance in intelligent tutoring systems ITS," *Int. J. Eng. Adv. Technol*., vol. 9, no. 3, pp. 2058–2066, Feb. 2020, doi: 10.35940/ijeat.C5169.029320.

[15] K. A. Assielou, C. Théodore, B. Tra, T. Lambert, and K. Daniel, "Emotional impact for predicting student performance in intelligent tutoring systems (ITS)," *Int. J. Adv. Comput. Sci. Appl*., vol. 11, no. 7, 2020, doi: 10.14569/IJACSA.2020.0110728.

[16] K. A. Assielou, C. T. Haba, and T. L. Kadjo, "Multi-relational and social-influence model for predicting student performance in intelligent tutoring systems (ITS)," *Int. J. Eng. Adv. Technol*., vol. 9, no. 3, pp. 2058–2066, 2020, doi: 10.35940/ijeat.c5169.029320.

[17] K. A. Assielou, C. T. Haba, T. L. Kadjo, B. T. Goore, and K. D. Yao, "A new approach to modelling students' socio-emotional attributes to predict their performance in intelligent tutoring systems," *J. Educ. e-Learning Res*., vol. 8, no. 3, pp. 340–348, Sep. 2021, doi: 10.20448/journal.509.2021.83.340.348.

[18] P. Nedungadi and T. K. Smruthy, "Personalized multi-relational matrix factorization model for predicting student performance," 2016, pp. 163–172.

[19] T.-N. Nguyen, N.-T. Mai, and H.-H. Nguyen, "An approach for multi-relational data context in recommender systems," *Asian Conference on Intelligent Information and Database Systems*, Springer, Cham, 2017.

[20] N. Thai-Nghe, M. Nhut-Tu, and H.-H. Nguyen, "An approach for multi-relational data context in recommender systems," 2017, pp. 709–720.

[21] D. T. G. Takács, I. Pilászy, and B. Németh, "On the gravity recommendation system," *Processing*, vol. 2007, pp. 22–30, 2007.

[22] N. Thai-Nghe and L. Schmidt-Thieme, "Multi-relational factorization models for student modeling in intelligent tutoring systems," in *Proc. - 2015 IEEE Int. Conf. Knowl. Syst. Eng. KSE 2015*, no. October, pp. 61–66, 2016, doi: 10.1109/KSE.2015.9.

[23] A. Villanueva *et al*., "Towards modeling of human skilling for electrical circuitry using augmented reality applications," *Int. J. Educ. Technol. High. Educ*., vol. 18, no. 1, p. 39, Dec. 2021, doi: 10.1186/s41239-021-00268-9.

[24] K. K. Tatsuoka, *Cognitive Assessment: An Introduction to the Rule Space Method*, Routledge, 2009.




**Denon Arthur Richmond Gono** got a degree in computer engineering. His current research focuses on pedagogical scenarization and student modeling in learning environments. He is currently a PhD student at the Polytechnic Doctoral School of INP-HB in Yamoussoukro and member of the Laboratory of Research in Computer Science and Telecommunications (LARIT). He is also an assistant professor at the National Institute of Technical and Vocational Education (Côte d'Ivoire).

**Bi Tra Goore** is currently a professor at INP-HB (Institut National Polytechnique Houphouët-Boigny of Yamoussoukro - Côte d'Ivoire). He spent several years teaching computer science and networks. He is the director of the Computer Science and Telecommunications Research Laboratory (LARIT). His research interests include the design, modeling and performance evaluation of wireless sensor networks, the design, modeling and performance evaluation of telecommunication networks and systems.

**Yves Tiecoura** did his master's degree and doctorate in Ukraine. He is an assistant professor (CAMES) at the Department of Electrical and Electronic Engineering of the Institut National Polytechnique Felix Houphouet Boigny (INP HB) in Yamoussoukro. He is a member of the Laboratory of Research in Computer Science and Telecommunications (LARIT). His field of specialization is industrial computer science.

**Kouamé Abel Assielou** is a master's degree in computer engineering from the University Nangui Abrogoua (Côte d'Ivoire), an engineering degree in networks and telecommunications from the BNETD (National Bureau of Technical Studies and Development) and a PhD from the INP-HB (Institut National Polytechnique Félix-Houphouët Boigny de Yamoussoukro) where he is currently working as an assistant professor. He is a member of the Laboratory of Research in Computer Science and Telecommunications (LARIT). His current research focuses on the design and modeling of intelligent tutoring systems (ITS).