1 Literature review

1.1 Adoptive learning

“ Adaptive learning as an educational technology is a kind of scaffolding technique customized to help all stakeholders in an educational institution, teachers, students and school administrators” (Castañeda & Selwyn, 2018)

According to Ennouamani and Mahani, (2018) there are 3 main adaptive learning approaches. They are ;

- Macro-Adaptive Approach - This approach allows the user to move between courses at an adapted rate. It also considers the learning objectives and cognitive and intellectual characteristics. The instructor has to initiate the narrative.
- Aptitude-Treatment Interaction (ATI) Approach - This approach identifies the learner's aptitude and then alters the course of action to improve the learner's abilities. These systems can be used to develop Intelligent Tutoring Systems by generating learning materials suited to individual learner's capabilities.
- Micro-Adaptive Approach - This approach analyzes the learner and understands the learner's requirement or knowledge gap. It is a more dynamic system that considers real-time characteristics of the learners.

This study focuses on building Aptitude-Treatment Interaction (ATI) Approach using deep learning. The ATI approach emphasizes the user's control over the learning process. Studies have shown that the success of self-control in learning depends on the learner's abilities, suggesting that it may be beneficial to limit control for students with low prior knowledge and enhance it for high-performing students. It introduces three levels of control: complete independence, partial control within task scenarios, and fixed tasks with controlled pace.

Intelligent Tutoring Systems (ITS) utilize the ATI approach to detect users' skills. ITS implementation is based on adaptive e-learning system architecture, comprising the learner model and domain model. An adaptation model is used to generate and present adapted materials to each learner. This approach is also applied in adaptive hypermedia systems, where the goal is to design learning solutions that integrate hypermedia content in ITS to tailor it to individual learner profiles

1.2 Knowledge tracing

Human teachers can measure students' level of understanding and take necessary actions to fill the gaps. In the computer base teaching era, machines must learn the students' degree of understanding and take action to fill the knowledge gap. Abdelrahman, Wang, and Nunes (2023) Recognize this process as **Knowledge Tracing (KT)** . These KTs are widely used in Massive Open Online Courses (MOOCs), Intelligent Tutoring Systems (ITS), educational games, and adaptive learning platforms. However, capturing student knowledge level is not easy because questions can require multiple skills, dependency among skills, and forgetting or decaying knowledge over time. Since John R. Anderson introduced knowledge tracing in 1986, researchers have attempted to develop many machine-learning models to solve KT.

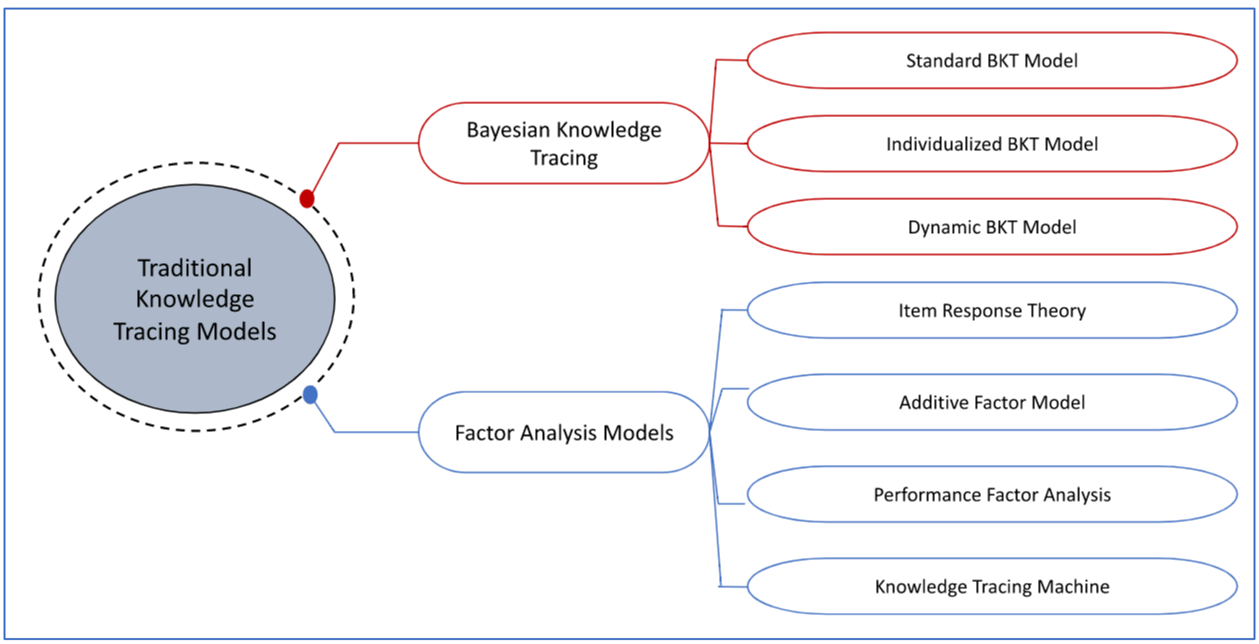

Figure 1.1- Types of Traditional Knowledge Tracing

1.3 Bayesian Knowledge Tracing

First generation of Traditional KT models were based on Bayesian Knowledge Tracing (BKT). Bayesian Knowledge Tracing (BKT) is inspired by mastery learning, which assumes that all students can achieve mastery of a skill under two conditions:

- Knowledge is organized as a hierarchy of skills, and
- Learning experiences are structured to ensure mastery of lower-level skills before moving to higher-level ones.

BKT models typically employ probabilistic graphical models like Hidden Markov Model and Bayesian Belief Network to track students' evolving knowledge states as they practice skills.

First BKT model was developed by Corbett and Anderson in 1994 (Albert T. Corbertt & John R Anderson, 1994). It considered two statues of student learned or unlearned. This model assumed that students do not forget what they mastered. . But it considers the probability that students may guess the answer p(G) or mistakenly select the wrong answer (slip) p(S).

This model consider as the standard BKT model. It has four parameters.

Table 1.1 - Bayesian Knowledge Tracing Formula Parameters

Parameter	Description
p(L0)	Probability of skill mastery by a student before learning
p(T)	Probability of transition from an unlearned state to a learned state
p(S)	Probability of slipping by a student in a learned state
p(G)	Probability of guessing correctly by a student in an unlearned state

At each time step n ≥ 1, the model estimates the probability p(Ln) of skill mastery by a student by

Equation 1.1- Bayesian Knowledge Tracing Formula

$$p(L_n) = Posterior(L_{n-1}) + (1 - Posterior(L_{n-1})) * p(T)$$

$$Posterior(L_{n-1}) = \begin{cases} \dfrac{p(L_{n-1}) * (1 - p(S))}{p(L_{n-1}) * (1 - p(S)) + (1 - p(L_{n-1})) * p(G)} & \text{if the } n\text{-th attempt is } \textit{correct}; \\ \dfrac{p(L_{n-1}) * p(S)}{p(L_{n-1}) * p(S) + (1 - p(L_{n-1})) * (1 - p(G))} & \text{otherwise.} \end{cases}$$

1.4 Factor analysis models

Factor analysis models are the second branch of traditional knowledge tracing methods. It plays a vital role in measuring assessments. Factor analysis models are based on Item Response Theory (IRT). This study use a data set from a commercial adaptive learning system that measures students proficiency using IRT. Our data set assed using a modified version of IRT which has a memory of previous performance and it helps to reach the mastery level based on the students adaptive rate.

Item Response Theory is a psychometrics method, which means it is statistical framework to analyze and understand the properties of individual test items/questions and the performance of test-takers on each item. It is introduced by F. M. Lord et al., 1968. IRT performance as a logistic function.

According to IRT every question has a degree of difficulty and student has a level of ability. Below equation is the basic form of ITR. p_{ij} is the probability of student *i* answering correctly to the question *j*. a_i is the ability of student *i* and b_j is the difficulty of the question *j*.

Equation 1.2 - Item Response Theory Formula

$$p_{ij} = \frac{e^{a_i - b_j}}{1 + e^{a_i - b_j}}$$

Assumptions in IRT;

- Probability of student correctly answering a question model as an item response function
- Item response function monotonically increase with respect to the ability of the student
- Questions are conditionally independent.

1.5 Deep knowledge tracing and Graph neural network

Piech et al., (2015) Lead the **Deep Knowledge Tracing (DKT)** introducing deep knowledge tracing . DKT mainly uses deep learning to predict students' ability to answer a question correctly. DKT models have outperformed traditional knowledge tracing models. From the machine learning perspective knowledge tracing is sequence modeling task. It try to predict the next state of the

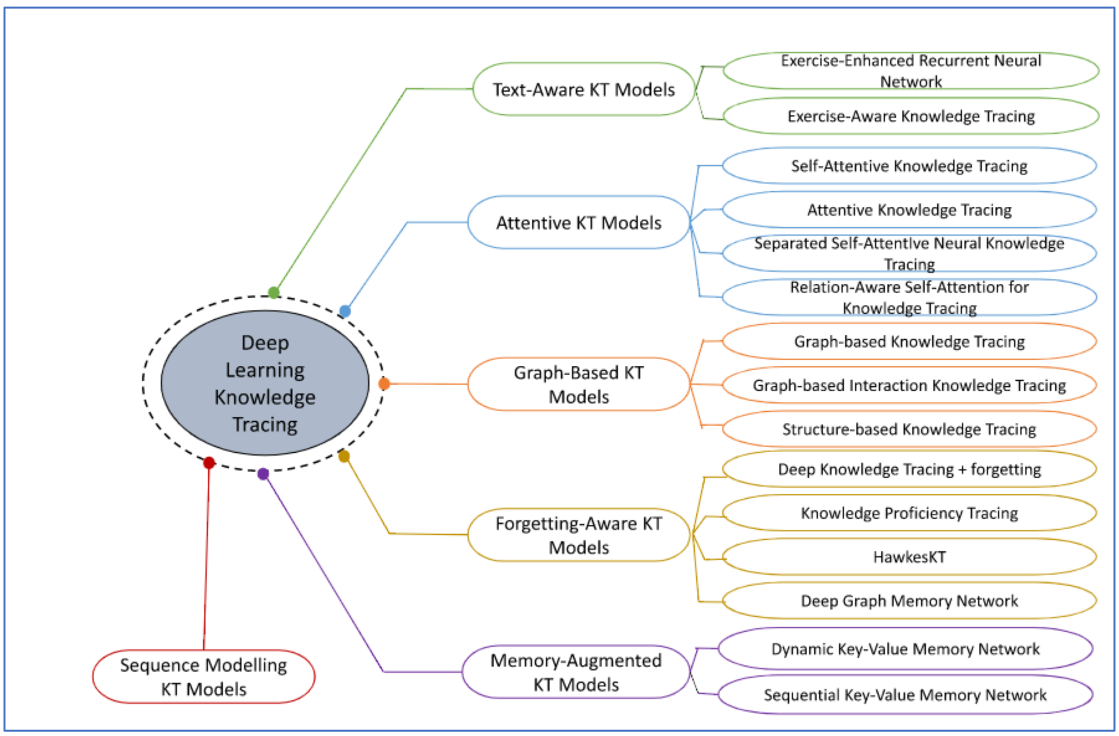

Figure 1.2 - Deep Learning Knowledge Tracing Methods

sequence (students ability to answer the next question given the previous questions and performance). Hence deep learning models have use Recurrent Neural Networks (RNN) or Long Short Memory (LSTM) to model the sequence. There are many branches under DKT.

Despite its encouraging performance DKT models had some drawbacks. They cannot predict the outcome of questions related to multiple skills. Also, it can not model the connection between multiple skills. It also assumes that all the questions are related to each other with the same probability which is not likely to happen all the time. Researchers try to overcome these limitations using Extended-deep knowledge tracing, which introduces by (Piech et al., 2015). They added additional student features such as previous knowledge, question answering rates and time spent on learning and practice; and, exercise features, such as textual information, question difficulty, skill hierarchies and skill dependencies.

But these limitations successfully overcome by Graph based Knowledge Tracing models. They can integrate the relationship between knowledge concepts and questions. Nakagawa et al., 2019 introduced Graph based knowledge tracing. They present knowledge concepts by nodes and relationships between them using edges. They formulated the problems as time series classification problem at node level.

According to (Abdelrahman et al., 2023) there are three main graph-based KT models. They are

- graph-based knowledge tracing
- graph-based interaction knowledge tracing
- structure-based knowledge tracing (SBKT)

This research leans toward structure-based knowledge tracing as we use knowledge graphs representing relationships between knowledge concepts (KC/learning objective (LO) as per our data set).

Tong et al.,(2020) introduced the structure-based knowledge tracing method. They have tried to solve two main challenges in this paper. They are the temporal impact of exercise sequence and the spatial impact of the knowledge structure or knowledge graph. In order to solve these challenges, they have introduced structure-based knowledge tracing(SBKT). SBKT can simultaneously model the temporal and spatial impacts.

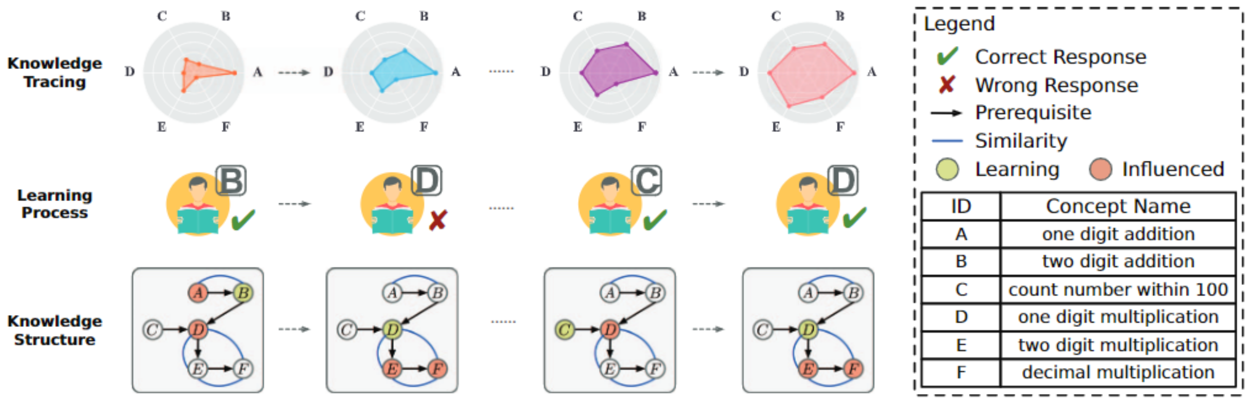

Figure 1.3 structure based knowledge tracing (Tong et al., 2020)

Figure 1.3 depict sequence of exercises related to one knowledge structure. Under this structurer there are connected concepts. They are either prerequisites or similar concepts. As the student proceed with the question students knowledge statues of each concept change. It is shown in the radar map in the top. Changes in radar map shows the temporal impact of the students' knowledge statues and knowledge structures shows how responses impact the learning concept and related(influenced) concepts, which is the spatial impact.

1.5.1 Graph Neural Network

The rapid development of internet technology and web applications has led to a vast amount of data being generated on the internet, which can be used to create valuable knowledge. Such knowledge lead to create knowledge graphs. Graph Neural Network (GNN) created to learn from such knowledge graphs and predict the unknown. GNN are a class of deep learning methods designed to perform inference on data described by graphs. They are neural networks that can be directly applied to graphs, allowing for node-level, edge-level, and graph-level prediction tasks (Ye et al., 2022).

The message parsing process is what allows GNNs to learn from the structure of the graph. By sending messages to each other, the nodes in the graph are able to share information about their local neighborhoods. This information can then be used to update the nodes' states, which in turn can be used to make predictions about the graph. There are a variety of different message parsing functions that can be used in GNNs. The choice of message parsing function depends on the specific task that the GNN is being used for. For example, if the GNN is being used to predict the relationship between two entities, then the message parsing function might be designed to extract features from the entities and their relationships (Serra & Niepert, 2023).

1.6 Leaners characteristics

Hemmler and Ifenthaler, (2022) have identified internal and external indicators of the learning context for supporting adaptive learning. Based on the authors internal dimensions, Past performance is a one dimension that support toward adaptive learning. It can be measure through previous grades, rank, previous experience with the course content, prior credits and course repetition. All these indicators are included in our data set. Additionally under skills and abilities dimension, prior knowledge indicator also captured in our data set. In contrary there are many other dimensions such as demographics, learning approach, emotions, perception towards teacher/course and etc. Hence our study limited only to student performance and skill/abilities dimension when analyzing learners characteristics in an adaptive learning environment.

Afini Normadhi et al.,(2019) summarize learners personal traits in 3 main domains and the relevant sub domains.

- Cognition – learning style /cognitive style/ prior knowledge/ personality type/thinking process/working memory capacity.
- Affective – emotions/ mental state/ engagement
- Behavior/psychomotor – cognitive abilities/ performance

Our study based on performance under Behavior/psychomotor and prior knowledge under cognition.

Authors conclude most of the adaptive learning environments build on personal traits under cognitive learning domain. Most frequently used personal trait identification method is computer based detection using machine learning (majority) , without machine learning or hybrid approach. Authors mentioned most of the research work suffer with small sample size which address in our study. And our work intend to use knowledge graph based approach which was not used mention in (Afini Normadhi et al., 2019) literature review from 2007-2017.

Hsu,(2012) developed Learning Effort Curve Mode using dynamic real-time based learning effort quantification technique (related work from the same author). This author has used learning style, learning efficiency and self-efficacy as learner characteristics. In the evaluation author has grouped 125 students in to 16 groups and measured Learning Effort Curve Mode. Author has found, despite the learning style or characteristics, descending learning effort leads to ascending learning performance for high learning efficacy groups . Similarly ascending learning effort leads descending learning performance low learning efficacy groups.

1.7 Recommendation system

Rule-based filtering systems rely on manually or automatically generated decision rules that are used to recommend items to users. Content-based filtering systems recommend items that are considered sufficiently similar to the content descriptions in the user profile. Collaborative filtering systems, also referred to as social filtering, match the rating of a current user for items with those of similar users in order to produce recommendations for items not yet rated or seen (Duval et al., 2007)

1.7.1 Study material recommendation

Duval, Klamma and Wolpers, (2007) developed an advance recommendation engine to recommend links to students in an E-learning platform. Regular recommendation engines, consider all the users logs at once to recommend links using sequential pattern mining algorithms. These authors have clustered users using k-means clustering algorithm (2-5 clusters) considering number of pages visited and the average knowledge obtained from these pages. Then they have applied AprioriAll, GSP and PrefixSpan sequential pattern mining algorithms for each cluster to generate recommendation rules. This new approach have generated similar or more rules for the same support and with high confidence compared to using all user data at once. As per the conclusions, GSP and PrefixSpan algorithms have shown better slightly better results when there are 2 or 3 clusters. In our approach we can generate 2 or 3 clusters to identify similar students. These authors haven’t consider the learning objectives but students navigation through the web site. Our work can also consider the number of questions and instruction materials referred and the student progress in the learn path (similar to average knowledge) as features for the clustering algorithm.

Our data set do not contain students activity log but students performance in relation to learning objectives. And the due graph nature of our data set make it more complex to analyze.

Borges and Stiubiener, (2014) developed a recommendation system to suggest learning materials to students based on the learning style of the students and the relevant learning objectives. Authors have clustered the students based on their learning style, they have identified 6 learning styles based on input , perception and process (Richard Felder, 2002), and how different learning materials associated with the learning style. Then utility function developed to measure the distance between learning objectives and learning style(LS) using Manhattan distance. Utility function range from 0 to 6, 0 indicate no difference between LO and LS. 6 indicate LO and LS is totally different from each other. Based on utility function results they and LS they suggest the learning materials. They have tested this system with 28 students and 362 recommendations, 89% of the students are satisfied with the results. In their research , they have not considered the students performance and applied for a small student group. Contrary in our study we consider students performance history and student performance after referring the learning materials. Our study based on large pool of students. Additionally we map LOs with knowledge graphs and how student performance related to each LO.

Reference

Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge Tracing: A Survey. *ACM Computing Surveys, 55*(11), 1–37. https://doi.org/10.1145/3569576

Afini Normadhi, N. B., Shuib, L., Md Nasir, H. N., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. *Computers and Education*, *130*, 168–190. https://doi.org/10.1016/j.compedu.2018.11.005

Albert T. Corbertt, & John R Anderson. (1994). Knowledge Tracing : Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*.

Borges, G., & Stiubiener, I. (2014). Recommending Learning Objects Based on Utility and Learning Style. *Opening IEEE-Innovations and Internationalization in Engineering Education* .

Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. In *International Journal of Educational Technology in Higher Education* (Vol. 15, Issue 1). Springer Netherlands. https://doi.org/10.1186/s41239-018-0109-y

Duval, E., Klamma, R., & Wolpers, M. (2007). LNCS 4753 - Personalized Links Recommendation Based on Data Mining in Adaptive Educational Hypermedia Systems. In *LNCS* (Vol. 4753).

Ennouamani, S., & Mahani, Z. (2018). An overview of adaptive e-learning systems. *2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017*, *2018-January*, 342–347. https://doi.org/10.1109/INTELCIS.2017.8260060

F. M. Lord, M. R. Novick, & Allan Birnbaum. (1968). *SOME LATENT TRAIT MODELS*.

Hemmler, Y. M., & Ifenthaler, D. (2022). Indicators of the Learning Context for Supporting Personalized and Adaptive Learning Environments. *Proceedings - 2022 International Conference on Advanced Learning Technologies, ICALT 2022*, 61–65. https://doi.org/10.1109/ICALT55010.2022.00026

Hsu, P.-S. (2012). LEARNER CHARACTERISTIC BASED LEARNING EFFORT CURVE MODE: THE CORE MECHANISM ON DEVELOPING PERSONALIZED ADAPTIVE E-LEARNING PLATFORM. In *TOJET: The Turkish Online Journal of Educational Technology* (Vol. 11, Issue 4).

Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019). *GRAPH-BASED KNOWLEDGE TRACING: MODELING STUDENT PROFICIENCY USING GRAPH NEURAL NET-WORK*.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., Sohl-Dickstein, J., University, S., & Academy, K. (2015). *Deep Knowledge Tracing*.

Richard Felder, by M. (2002). *LEARNING AND TEACHING STYLES IN ENGINEERING EDUCATION*. http://www.ncsu.edu/felder-public/ILSpage.html

Serra, G., & Niepert, M. (2023). *L2XGNN: Learning to Explain Graph Neural Networks*.

Tong, S., Liu, Q., Huang, W., Huang, Z., Chen, E., Liu, C., Ma, H., & Wang, S. (2020). Structure-based knowledge tracing: An influence propagation view. *Proceedings - IEEE International Conference on Data Mining, ICDM*, *2020-November*, 541–550. https://doi.org/10.1109/ICDM50108.2020.00063

Ye, Z., Kumar, Y. J., Sing, G. O., Song, F., & Wang, J. (2022). A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs. *IEEE Access*, *10*, 75729–75741. https://doi.org/10.1109/ACCESS.2022.3191784