Article

Predicting Task Planning Ability for Learners Engaged in Searching as Learning Based on Tree-Structured Long Short-Term Memory Networks

Pengfei Li [1], Shaoyu Dong [2,*], Yin Zhang [3] and Bin Zhang [3]

1   School of Computer Science and Engineering, Northeastern University, Shenyang 110167, China; leepengfei.neu@gmail.com
2   School of Business Administration, Northeastern University, Shenyang 110167, China
3   Software College, Northeastern University, Shenyang 110167, China; zhang-yin@mail.neu.edu.cn (Y.Z.); zhang-bin@mail.neu.edu.cn (B.Z.)
*   Correspondence: sydong_91@163.com

Abstract: The growing utilization of web-based search engines for learning purposes has led to increased studies on searching as learning (SAL). In order to achieve the desired learning outcomes, web learners have to carefully plan their learning objectives. Previous SAL research has proposed the significant influence of task planning quality on learning outcomes. Therefore, accurately predicting web-based learners' task planning abilities, particularly in the context of SAL, is of paramount importance for both web-based search engines and recommendation systems. To solve this problem, this paper proposes a method for predicting the ability of task planning for web learners. Specifically, we first introduced a tree-based representation method to capture how learners plan their learning tasks. Subsequently, we proposed a method based on the deep learning technique to accurately predict the SAL task planning ability for web learners. Experimental results indicate that, compared to baseline approaches, our proposed method can provide a more effective representation of learners' task planning and deliver more accurate predictions of learners' task planning abilities in SAL.

Keywords: searching as learning; learning ability; HCDP; Tree-Structured Long Short-Term Memory Networks; user analysis; task planning

1. Introduction

In recent years, the advent of web-based search engines has revolutionized the way people access information. These ubiquitous tools are extensively employed, not only for informational queries but also increasingly for learning purposes [1,2]. Recognizing the potential of web-based search engines as valuable learning aids, researchers have focused on searching as learning (SAL), utilizing web-based search engines as a means to acquire knowledge and support learning processes and conceptualizing searching activities as learning activities [3,4].

Unlike the traditional field of information retrieval, which primarily views searching as a tool for information acquisition, studies on SAL place greater emphasis on the learning process that learners engage in through web-based searching. Building upon this perspective, research in the domain of SAL focuses on the role of search systems in directly facilitating human learning [5]. This area of study goes beyond mere information retrieval to emphasize examining the effects, implications, and results derived from utilizing search systems in the context of educational processes. From a perspective of information retrieval, SAL research shifts the focus from the relevance of individual search results to supporting the learning process itself [6]. From an educational perspective, SAL research concentrates on deeply understanding how learners use search engines to meet their learning needs and how optimization can enhance learning outcomes [2].

Studies have proposed that SAL combined with thoughtful task planning can lead to enhanced learning outcomes [7,8]. At the beginning of the SAL process, learners often only possess a vague understanding of the learning object, meaning their knowledge structures are insufficient to precisely articulate what they seek to learn. During the SAL process, learners are required to continually refine their learning tasks and retrieve relevant information from web-based search engine results, progressively constructing and refining their knowledge structures. This process involves the generation of queries, the evaluation of search results, and iterative adjustments and refinements of knowledge structures [5,9].

Understanding and predicting learners' task planning ability in SAL are crucial for web-based search engine providers, recommendation systems, and educators [6,10]. By comprehending learners' planning abilities, web-based search engines and recommendation systems can provide targeted guidance, suggest relevant learning resources, and optimize search results to facilitate effective learning [11]. Additionally, educators and instructional designers can utilize these insights to tailor instruction, provide appropriate scaffolding, and design interventions aimed at improving learners' planning skills, ultimately fostering metacognitive awareness and self-regulated learning [12].

To address the challenge of predicting learners' task planning ability, this paper proposes a novel method that leverages the Hierarchical Clustering Algorithm Based on Density Peaks (HCDP) model and the Tree-Structured Long Short-Term Memory Networks (Tree-LSTM) algorithm. The HCDP model is employed to capture and represent the hierarchical relationships among learning activities and learning subtasks. By modeling the learning process in this way, we can effectively capture the nuances of task planning in SAL. The Tree-LSTM algorithm is utilized to predict learners' SAL task planning ability based on the extracted features from the tree structure.

The experimental results demonstrate the efficacy of our proposed method in effectively predicting learners' task planning ability within the context of SAL. Furthermore, the key features extracted from the tree structure serve as reliable indicators of learners' planning ability, providing valuable insights for web-based search engines, recommendation systems, and instructional designers.

Overall, this paper contributes to the field of information retrieval and learning by offering a methodological approach to predict learners' task planning ability in the context of SAL. The findings hold implications for web-based search engine providers, recommendation systems, and educational practitioners. For search engine designers, our study aids in developing learner-focused search interfaces by understanding SAL task planning, leading to enhanced personalization and efficiency. For educational practice, our research informs educators about learner challenges in SAL, enabling more effective learning experience design and targeted support.

2. Related Works

2.1. Predictive Models for Learners' Abilities

The predictive modeling of learners' abilities has gained significant interest, especially due to its potential in customizing learning environments for individual learners, thereby optimizing the learning process.

Numerous models have been proposed that utilize Machine Learning (ML) and Artificial Intelligence (AI) algorithms to predict abilities. For example, Thai-Nghe et al. [13] introduced a method to predict student performance based on past interactions using collaborative filtering and matrix factorization techniques. Similarly, Márquez-Vera et al. [14] employed decision trees, Naive Bayes, and k-nearest neighbors to anticipate student dropouts in online courses. Liu et al. [15] propose a two-stage framework to predict the cognitive level of the learner. Agrawal et al. [16] pointed out that learning ability can be estimated by administering a test designed using modern practices such as those based on Item Response Theory (IRT). Bockmon et al. [17] conducted a comprehensive study on the predictive modeling of students' introductory programming abilities at the end of the semester. To achieve this, they employed a multinomial logistic regression approach, aiming to de-

velop a robust model that could effectively forecast students' performance in programming tasks. This model's sophistication lies in its ability to handle multiple predictor variables and their interactions, offering a nuanced understanding of student performance. Furthermore, the research delved into the relationship between various factors, such as prior programming abilities, spatial skills, socioeconomic status, and students' attitudes toward computing, in order to determine their influence on the final programming outcomes. In sum, this comprehensive analysis provides a foundation for developing targeted educational strategies that can significantly improve student outcomes in programming and related technical disciplines.

The advancements in predictive modeling underscore the importance of understanding learners' abilities, which is a crucial aspect of Searching as Learning (SAL). This understanding aids in the development of more effective web-based learning tools and strategies.

2.2. Searching as Learning

The growing utilization of web-based search engines as tools for learning has attracted considerable attention from researchers. Studies have explored the impact of web-based search engine features, such as query formulation assistance, result evaluation techniques, and personalized recommendations, on learning outcomes [7,18]. These investigations highlight the significance of effective web-based search engine usage in supporting the learning process [19,20].

For instance, query formulation assistance helps learners in generating effective search queries, enabling them to retrieve relevant and accurate information [6,21,22]. Result evaluation techniques aid learners in critically assessing the credibility, relevance, and reliability of search results, enabling them to make informed decisions regarding the information they encounter during their learning process [23].

SAL studies use searching as a part of the learning process and aim to explore the integration of web-based search engine utilization and web learning to improve learning outcomes [6,24,25]. Similar SAL studies conceptualize searching as an integral component of the learning process and underscore the significance of search in enhancing learning outcomes [6,26,27]. These studies emphasize the importance of task planning quality, including query formulation and result evaluation, in achieving desired learning outcomes.

3. Data Collection and Labeling

In this section, we discuss the dataset employed in our experiments that was procured from the University Writing Program (UWP) courses at Northeastern University. Further, we discuss how we achieved capability labeling for task planning in the SAL context.

3.1. Data Collection

In this section, we begin by detailing the SAL dataset utilized in this study, collected from learners enrolled in the UWP course at Northeastern University. The data collection methodology has been described in our previous work [6]. Here, we briefly outline the types of SAL data captured for each learner:

(1) Search logs. We recorded learners' search activities with web-based search engines by developing a Firefox browser plug-in. Specifically, for a learner, we recorded searching activities such as their issued search queries, clicking on URLs, and reading duration times.
(2) Search results. We recorded search results after each issued search query.
(3) Learning outcomes. We recorded programming snapshots for each learner during compilation.

In the initial five weeks of our study, we systematically introduced tasks with an incrementally increasing number of subtasks. The distribution of these subtasks within the assignments is illustrated in Figure 1. Research conducted in the domain of SAL has suggested that the act of searching within SAL can be perceived as a sequence of activities

with the purpose of learning [28]. Although we cannot directly observe, we can predict their learning state by analyzing their search activities and learning outcomes.
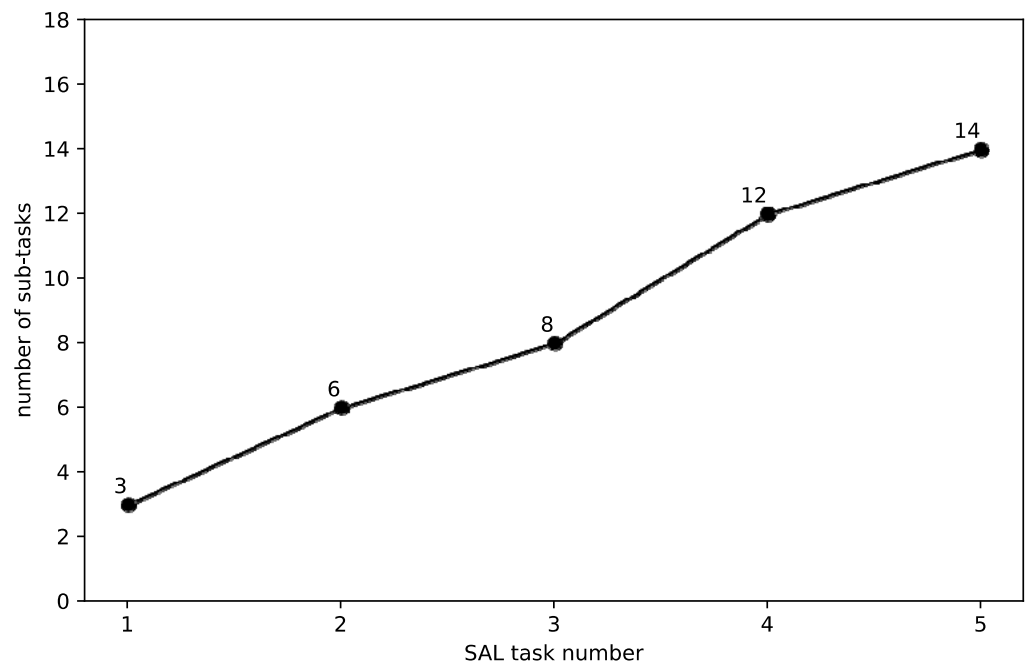

Figure 1. The distribution of subtasks in the learning assignments.

3.2. Data Labeling

In this study, six researchers specializing in the field of SAL from Northeastern University participated in the labeling process. These experts included two associate professors, two doctoral candidates, and two master's degree learners. Each participant was tasked with evaluating the SAL abilities of learners based on the dataset we collected. This evaluation involved reviewing the collected data in conjunction with the scores of corresponding assignments within the course curriculum. The average scores from five distinct assignments were ultimately employed as the annotated indicators for gauging the SAL abilities of the learners.

To facilitate the labeling of SAL task planning abilities, it is first necessary for each participant to analyze the learning process of learners in order to understand how these learners decompose their learning tasks. To ensure the validity of manual annotations, we required participants to answer specific questions given different types of interaction behaviors. This ensures a comprehensive understanding of the learner's learning process during analysis, as illustrated in Table 1. Additionally, participants can also refer to the learner's phased learning outcomes and final grades for labeling.

The statistical outcomes of the capability labeling for task planning in the SAL context is illustrated in Figure 2. For the purpose of this research, the manually annotated SAL task planning abilities are classified into five distinct levels, ranging from 1 to 5. A score of 1 represents the lowest level of capability, while a score of 5 signifies the highest level of proficiency in SAL task planning.

Table 1. Questions that need to be answered for different types of SAL behavior.

Issuing queries
1. Why did the learner issue this query?
2. Is this query related to the previously submitted queries?
3. What is the relationship between the results returned by this query and the results returned by previous queries?

Clicked on URLs
1. What learning object is the learner interested in?
2. Was this click event triggered by the most recent query?
3. Is the learner's learning objective the same as or related to the learning objective of the previously submitted queries?

Programming
1. Through which queries did the learner acquire his/her learning outcomes?
2. To achieve the learning outcomes, did the learner experience struggles or study unrelated content?
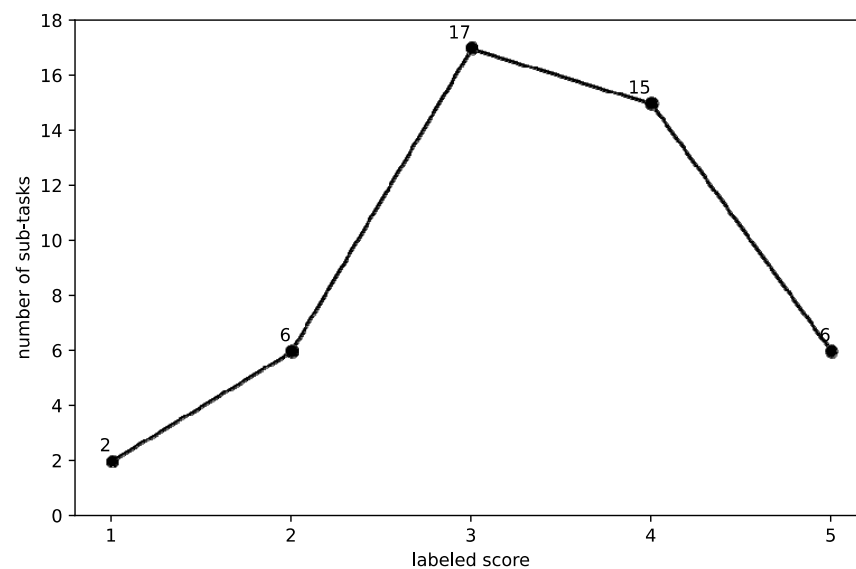
Figure 2. Results of SAL Task planning Capability Labeling.

4. Proposed Methodology

In this section, we will accomplish two primary objectives. First, we innovatively employ the HCDP method for constructing tree-like structures, effectively enabling the hierarchical representation of SAL task planning. Second, we introduce the use of the Tee-LSTM approach to facilitate the prediction of SAL task planning abilities.

4.1. Representation of Task Planning in Searching as Learning

In this section, we focus on the construction of a structured representation for task planning in SAL. While linear structures have been extensively employed for representing task planning, recent research has indicated that learning processes are often intricate search activities requiring learners to navigate among varying learning objectives and tasks [29]. Consequently, a linear structure proves insufficient for capturing the complexity inherent in a learner's task planning strategies.

Mehrotra et al. [30] substantiated the advantages of tree-structured representations in modeling search task planning. Moreover, current research has indicated that hierarchical clustering algorithms can effectively capture the subtask structure of learners' search tasks [6]. Accordingly, in the present study, we adopt a tree-structured approach to provide a more nuanced and effective representation of task planning in SAL.

To fully leverage the unique context-specific features of SAL and to provide a more accurate representation of a learner's task-based structural divisions, we introduce a novel method for SAL task partitioning based on HCDP. This advanced hierarchical clustering algorithm allows for capturing the intricacies of searching and learning interactions in SAL [31].

To accurately model the structure of a learner's task planning, we consolidate his/her SAL-related interactive activities prior to initiating the modeling process. We complete this based on the observation that learning activities are triggered by issuing search queries. Furthermore, what learners acquire is contingent upon the queries they submit. Hence, in constructing the structural representation of a learner's task planning, we employ queries as the nodes of the structure. While these nodes are represented by queries, it should be noted that they encapsulate not only the search queries themselves but also the subsequent learning that occurs as a result.

In traditional HCDP, the algorithmic framework is fundamentally structured around three core procedures: the computation of local densities, the construction of a hierarchical representation of the data, and the extraction of optimal clusters [32–34]. Given that our research objective specifically aims to establish a hierarchical architecture for task partitioning in SAL, our study focuses only on executing the initial two procedures.

HCDP employs k-nearest neighbors for the computation of local densities. The HCDP model computes the local density as follows [34]:

$$\rho_i = \max_{j \in knn(i)} dist(i, j) \tag{1}$$

where ρ_i is the $Node_i$'s k-nearest density, $dist(i, j)$ denotes the distance between $Node_i$ and $Node_j$.

For each $Node_i$, SAL establishes a connection to its nearest neighbor with higher density using edge weight φ. The computation for φ is as follows.

$$\varphi_i = \min_{j:\rho_i > \rho_j} dist(i, j) \tag{2}$$

Therefore, in our task of hierarchical representation for task planning, the focus is on being able to calculate $dist(i, j)$ by integrating features from SAL. To achieve this goal, we calculate $dist(i, j)$ from three dimensions: search, learning, and the connection between search and learning. We list the SAL features for calculat $dist(\cdot)$ that we employed in Table 2. The hierarchical clustering visualization of partial SAL data for a learner in the UWP dataset is shown in Figure 3.
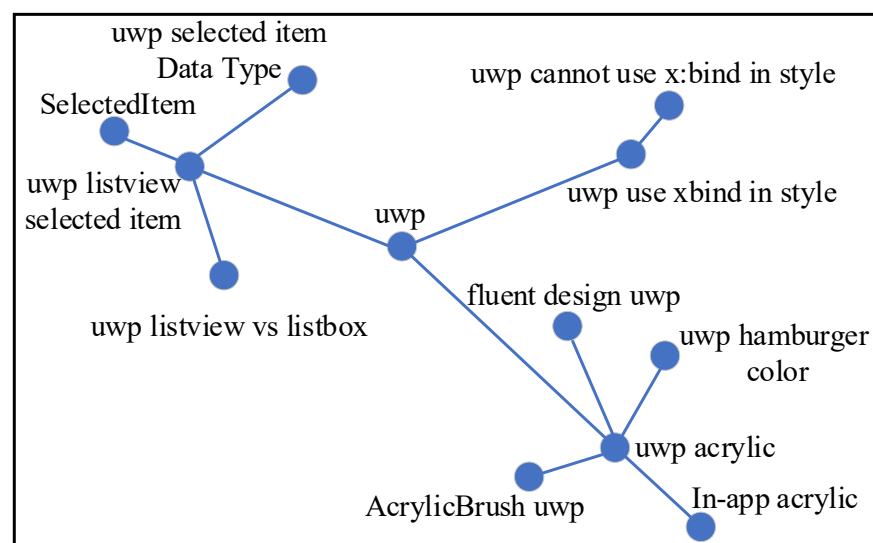

Figure 3. The hierarchical clustering visualization of partial SAL data for a learner.

Table 2. The SAL features for calculating $dist(\cdot)$.

Search-related features
1. Cosine distance between two sets of query terms.
2. Edit distance between two sets of query terms.
3. Jaccard distance between two sets of query terms.
4. The proportion of identical terms in two search queries.
5. Semantics distance between queries.
Features of the relationship between searching and learning
1. The average cosine distance between the web page links clicked after queries.
2. The average edit distance between the web page links clicked after queries.
3. Cosine distance between the sets of UWP terms contained in clicked links after two queries.
4. Cosine distance between the sets of UWP terms contained in the search results after two queries.
Learning-related features
1. Cosine distance between the sets of UWP classes contained in programming snapshots after two queries.
2. Edit distance between the sets of UWP classes contained in programming snapshots after two queries.
3. Semantic distance between two programming snapshots.

4.2. SAL Task Planning Ability Predicted Based on the Tree-LSTM Model

In this paper, we address the challenge by employing the Tree-LSTM model. A Tree-LSTM is a neural network architecture that extends the standard Long Short-Term Memory (LSTM) framework [35]. While standard LSTM models are designed to process sequential data, Tree-LSTM models are adapted to handle tree-structured data. This makes them particularly useful for tasks that involve hierarchical or nested structures, such as natural language sentences, computer programs, or chemical molecules [31]. Therefore, the Tree-LSTM model serves as an instrumental methodology, enabling a more nuanced understanding of hierarchical dependencies and thereby predicting task planning ability from tree hierarchical representation.

The key advantage of Tree-LSTMs lies in their ability to capture the hierarchical dependencies within tree-structured data [35,36]. This is particularly beneficial in educational contexts where learning tasks often involve layered concepts or stepwise procedures. For instance, in the realm of programming education, the Tree-LSTM model can effectively represent and analyze the structure of code, discerning the underlying logic and predicting potential errors or areas of improvement in student submissions [37,38].

In the context of our research framework, the Tree-LSTM model ingests a tree-structured representation encapsulating the complexities of the learning task as its input. The tree-structured data input represents the hierarchical organization of a learning task, capturing various elements such as the sequence of steps, dependencies among concepts, and the progression of learning objectives. The Tree-LSTM model then processes this input to generate an output in the form of a predictive assessment, quantifying a learner's abilities in task planning. Analogous to the conventional LSTM model, each unit in a Tree-LSTM architecture is equipped with input gates denoted as i_m, output gates symbolized by o_m, along with a memory cell c_m and a hidden state h_m. Unique to the Tree-LSTM model, the updating mechanism for these gate vectors and memory cells is conditioned upon the aggregated states of multiple child units, if present. Moreover, each Tree-LSTM unit is endowed with specialized forget gates $f_{m,k}$ for each child unit k [39,40]. This design intricacy enables the Tree-LSTM to serve as a robust framework for modeling hierarchical relationships, particularly valuable for tree-structure presentation of learning tasks.

Let $S(m)$ denote the subtree of m, and the transition equations of the Tree-LSTM model are as follows [41]:

$$\widetilde{h_m} = \sum\nolimits_{m \in S(j)} h_n \tag{3}$$

$$i_m = \sigma\left(W^{(i)}x_m + U^{(i)}h_m + b^{(i)}\right) \tag{4}$$

$$f_{m,k} = \sigma\left(W^{(f)}x_m + U^{(f)}h_m + b^{(f)}\right) \tag{5}$$

$$o_m = \sigma\left(W^{(o)}x_m + U^{(o)}h_m + b^{(o)}\right) \tag{6}$$

$$u_m = \tan h\left(W^{(u)}x_m + U^{(u)}h_m + b^{(u)}\right) \tag{7}$$

$$c_m = i_m \circ u_m + \sum_{k \in S(m)} f_{m,k} \circ c_k \tag{8}$$

$$h_m = o_m \circ \tan h(c_m) \tag{9}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid function, and $\circ$ denote the element-wise multiplication. In the Tree-LSTM model, the state of the composite nodes is derived from the states of the nodes, as illustrated in Figure 4.
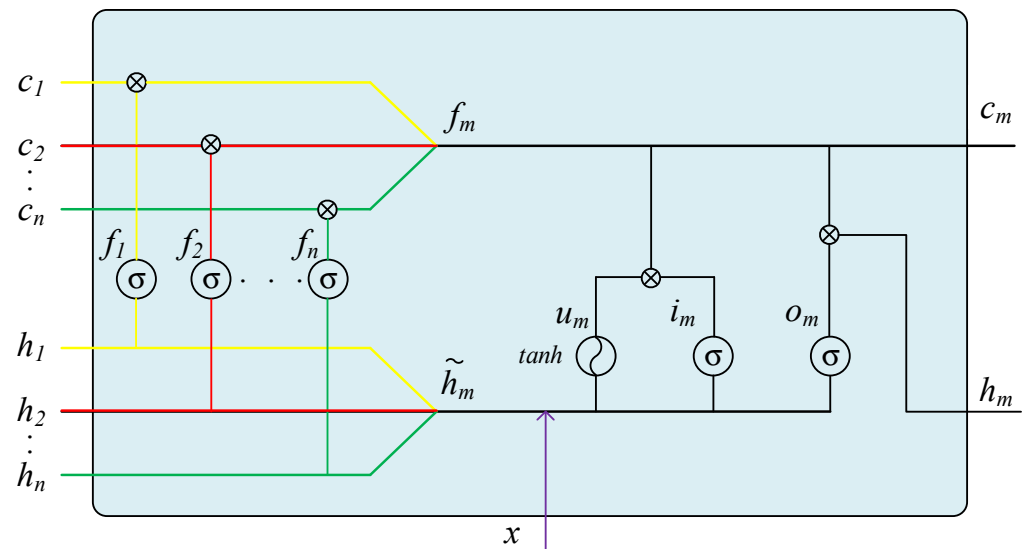

Figure 4. Tree-LSTM model for task planning ability prediction.

The inclusion of these additional gates and the unique updating mechanism enables the Tree-LSTM model to effectively capture and analyze the intricacies of hierarchical data, making it a powerful tool for modeling the dynamic nature of learning tasks. This advanced functionality positions the Tree-LSTM as an ideal framework for tasks that require an understanding of nested or sequential dependencies, such as predicting a learner's ability to plan and execute complex learning tasks.

5. Experiments

To assess the efficacy of our proposed methodology in forecasting learners' abilities in task planning within the SAL, we executed an array of experiments utilizing the Northeastern University UWP dataset as our empirical foundation. This section commences by detailing the experimental setup. Subsequently, we substantiate the merits of our approach by juxtaposing its performance metrics against those of established baseline algorithms.

5.1. Experimental Setup

To verify the performance of our proposed method, we commence by delineating the experimental setup. The design of our experiments employed the dataset gathered

from the UWP course at Northeastern University (China), and the manually labeled task planning abilities that we discussed in Section 3.2. To ensure the reliability of our findings, we employed stratified ten-fold cross-validation for dataset partitioning into training and testing subsets. The rationale behind utilizing stratified ten-fold cross-validation lies in its capacity to mitigate the introduction of potential biases and anomalous results, which may stem from imbalanced or skewed data distributions.

For comparative model analysis, our methodology underwent a two-phase evaluation. Initially, we compared our proposed method with state-of-the-art (SOAT) hierarchical clustering algorithms, thereby establishing the performance efficiency of the HCDP algorithm in the hierarchical representation within the SAL task planning. Subsequently, our framework was benchmarked against baseline predictive models for assessing the model's predictive accuracy.

5.2. Comparison with SOAT Hierarchical Clustering Methods Based on the UWP Dataset

In this section, we evaluate the advanced nature of our proposed methodology in the domain of learning task-structured representation through comparative experiments with SOAT methods. Selected methods for comparison include hierarchical clustering methods like Bayesian Hierarchical Clustering (BHC) [42], Min-Min-Roughness (MMR) [43], and Bayesian Rose Tree (BRT) [30]. A commonality between these methods and our proposed approach is their capability to construct hierarchical representations for learning task-planning. To ensure fairness and validity in the comparative analysis, all methods utilize SAL features consistent with those presented in Table 3 wherever possible. During the prediction phase, all of these hierarchical clustering methodologies employ the same Tree-LSTM model and undergo parameter optimization through identical procedures.

Table 3. The experimental results with hierarchical clustering methods.

Method	Precision	Recall	F1
BHC	0.717	0.701	0.709
MMR	0.782	0.73	0.755
BRT	0.82	0.805	0.812
Our method	0.889	0.825	0.856

As illustrated in Table 3, it is evident that the methodology proposed in this study demonstrates a superior performance over the baseline methods across multiple evaluation metrics. Specifically, the proposed approach surpasses the best-performing baseline method by approximately 7.1% in terms of average prediction accuracy for the UWP dataset. Notably, our method's performance exceeds that of the original BRT model, thereby substantiating the efficacy of the proposed model in the learning process. Moreover, among all methods, the BHC method exhibits the weakest predictive performance. This can be attributed to the fact that the binary tree structure is not congruent with the structural nuances of the learning process. In Table 4, the confusion matrix corresponding to the method we have proposed is delineated. This matrix effectively illustrates the performance of our methodology in terms of true positives, false positives, true negatives, and false negatives. Through this representation, we aim to provide a clear and comprehensive understanding of the accuracy, precision, recall, and specificity of our approach.

Table 4. The confusion matrix our method.

Method	TP	FP	TN	FN
Our method	104	13	91	22

Further, we conducted a comparative analysis of various algorithms' predictive capabilities across learning tasks with varying numbers of subtasks. As illustrated in Figure 5

(where the X-axis represents the number of subtasks in a learning task), with an increase in the number of subtasks, the prediction accuracy of the method proposed in this paper declines less compared to that of other baseline methods. When the number of tasks reaches 14 (which corresponds to the assignment with the most subtasks in this course), the difference in prediction accuracy between our proposed method and the best-performing baseline algorithm is at its maximum. In summary, as the number of subtasks increases, the performance advantage of our proposed algorithm becomes increasingly evident.
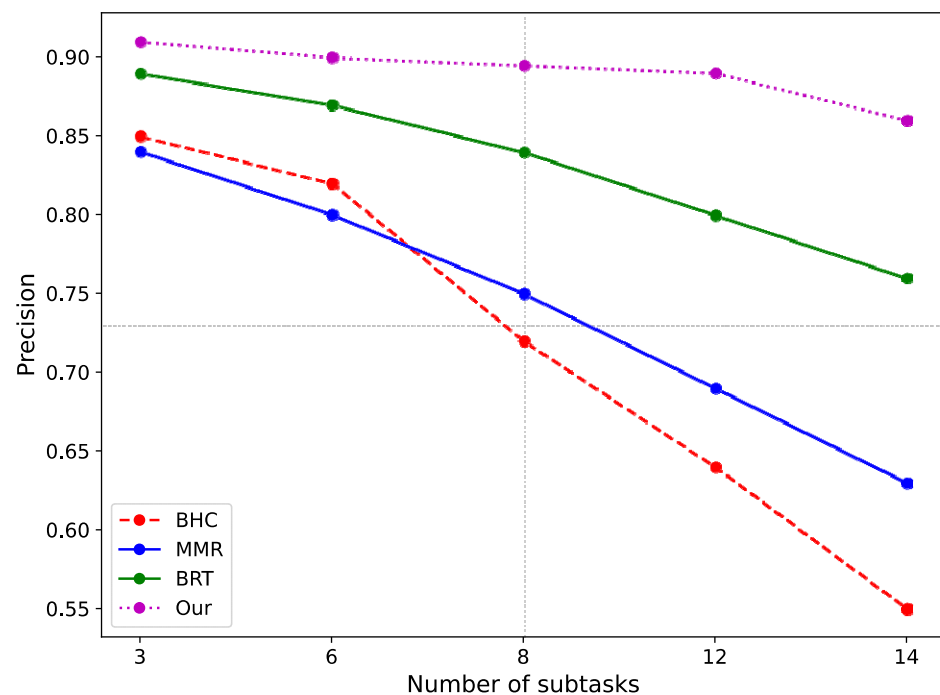

Figure 5. Comparison results for different numbers of subtasks.

5.3. Comparison with Predicative Methods

In this section, we evaluate the efficacy of our proposed methodology in the realm of task planning capability prediction by contrasting it with baseline approaches. The selected comparative methodologies include fundamental ML algorithms like Graph Neural Networks (GNNs) [44] and Recursive Neural Networks (RecNNs) [45]. These models were chosen for their capacity to accommodate tree-structured input data, thereby ensuring a level playing field for comparative analysis. The input to all of the models was constructed using HCDP, a preprocessing technique suited for SAL. During the training phase, parameter optimization was performed across all models to ensure performance.

Table 5 showcases a comprehensive evaluation of various algorithms, including our proposed methodology. The results elucidated in this table are a testament to the effectiveness of our technique. It is evident from the empirical data that our approach has a definitive edge over the baseline methodologies.

Table 5. The experimental results with baseline predicative methods.

Method	Precision	Recall	F1
GNN	0.843	0.824	0.833
RecNN	0.835	0.817	0.826
Our method	0.889	0.825	0.856

In the field of SAL, the precision and accuracy of predictions hold paramount significance. Given the complexities inherent to SAL, it is imperative for algorithms to adeptly

predict and optimize task planning. As delineated in Table 5, our methodology distinctly excels in this dimension. It not only assures enhanced accuracy but also emphasizes the salience of context-aware predictions within SAL. Multiple elements bolster the pre-eminence of our approach. Primarily, the strategy we introduced is congruent with the task planning architecture intrinsic to learners. Subsequently, the Tree-LSTM exhibits remarkable efficacy in modeling and predicting tree-structured data.

6. Conclusions

This research innovatively introduces a novel method for the accurate prediction of task planning abilities in the context of SAL. By utilizing the HCDP algorithm, we offer a hierarchical representation of the task planning for learners engaged in SAL. Leveraging the Tree-LSTM algorithm, we subsequently achieve precise predictive abilities for assessing task planning in SAL. Empirical validation, based on the UWP dataset, corroborates the effectiveness of our proposed approach.

For search engine designers, our research will assist web-based search engine designers in constructing learner profiles and in understanding how learners progressively complete their tasks in the context of SAL. Additionally, our findings will guide designers in creating more personalized and efficient search interfaces tailored for educational purposes. Moreover, our research can inform the optimization of query suggestions and the customization of result filtering based on learners' task planning abilities in SAL.

For educational practice, our research will significantly aid educational practitioners in designing more effective learning experiences. Specifically, it will help practitioners promptly identify and address the challenges and struggles learners may encounter, offering robust support in instructional design. Furthermore, this understanding will enable practitioners to provide targeted guidance and support, particularly for learners who struggle with planning and organizing learning tasks.

Future research avenues may encompass the analysis and understanding of various other abilities demonstrated by learners throughout the learning process. Additionally, the role of metacognition in influencing learning trajectories within SAL contexts warrants further investigation.

Author Contributions: Conceptualization, P.L. and B.Z.; Data curation, P.L., S.D. and Y.Z.; Formal analysis, P.L., B.Z. and Y.Z.; Funding acquisition, B.Z.; Investigation, P.L. and S.D.; Methodology, P.L. and Y.Z.; Resources, P.L. and S.D.; Software, P.L.; Validation, P.L., S.D., B.Z. and Y.Z.; Writing—original draft, P.L.; Writing—review & editing, P.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rahman, M.M.; Abdullah, N.A. A Personalized Group-Based Recommendation Approach for Web Search in E-Learning. *IEEE Access* **2018**, *6*, 34166–34178. [CrossRef]
2. Von Hoyer, J.; Hoppe, A.; Kammerer, Y.; Otto, C.; Pardi, G.; Rokicki, M.; Yu, R.; Dietze, S.; Ewerth, R.; Holtz, P. The Search as Learning Spaceship: Toward a Comprehensive Model of Psychological and Technological Facets of Search as Learning. *Front. Psychol.* **2022**, *13*, 827748. [CrossRef]
3. Zhang, P.; Soergel, D. Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *J. Inf. Sci.* **2016**, *42*, 59–78. [CrossRef]
4. Su, Y.-S.; Huang, C.S.; Ding, T.-J. Examining the Effects of MOOCs Learners' Social Searching Results on Learning Behaviors and Learning Outcomes. *Eurasia J. Math. Sci. Technol. Educ.* **2016**, *12*, 2517–2529. [CrossRef]

5. Hansen, P.; Rieh, S.Y. Editorial: Recent advances on searching as learning: An introduction to the special issue. *J. Inf. Sci.* **2016**, *42*, 3–6. [CrossRef]

6. Rieh, S.Y.; Collins-Thompson, K.; Hansen, P.; Lee, H.-J. Towards searching as a learning process: A review of current perspectives and future directions. *J. Inf. Sci.* **2016**, *42*, 19–34. [CrossRef]

7. Piech, C.; Sahami, M.; Koller, D.; Cooper, S.; Blikstein, P. Modeling how students learn to program. In Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12), Raleigh, NC, USA, 29 February–3 March 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 153–160. [CrossRef]

8. Li, P.; Zhang, B.; Zhang, Y. Extracting Searching as Learning Tasks Based on IBRT Approach. *Appl. Sci.* **2022**, *12*, 5879. [CrossRef]

9. Vakkari, P. Searching as learning: A systematization based on literature. *J. Inf. Sci.* **2016**, *42*, 7–18. [CrossRef]

10. Bhattacharya, N. LongSAL: A Longitudinal Search as Learning Study with University Students. In Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), Hamburg, Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023; p. 570. [CrossRef]

11. Liu, J. Deconstructing search tasks in interactive information retrieval: A systematic review of task dimensions and predictors. *Inf. Process. Manag.* **2021**, *58*, 3. [CrossRef]

12. Reynolds, R.B. Relationships among tasks, collaborative inquiry processes, inquiry resolutions, and knowledge outcomes in adolescents during guided discovery-based game design in school. *J. Inf. Sci.* **2016**, *42*, 35–58. [CrossRef]

13. Thai-Nghe, N.; Drumond, L.; Horváth, T.; Krohn-Grimberghe, A.; Nanopoulos, A.; Schmidt-Thieme, L. Factorization techniques for predicting student performance. In *Educational Recommender Systems and Technologies: Practices and Challenges*; Santos, O.C., Boticario, J.G., Eds.; IGI Global: Hershey, PA, USA, 2012; pp. 129–153.

14. Marquez-Vera, C.; Morales, C.R.; Soto, S.V. Predicting School Failure and Dropout by Using Data Mining Techniques. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2013**, *8*, 7–14. [CrossRef]

15. Liu, Y.; Liu, Q.; Wu, R.; Chen, E.; Su, Y.; Chen, Z.; Hu, G. Collaborative Learning Team Formation: A Cognitive Modeling Perspective. In *Database Systems for Advanced Applications, Proceedings of the 21st International Conference, DASFAA 2016, Dallas, TX, USA, 16–19 April 2016*; Navathe, S., Wu, W., Shekhar, S., Du, X., Wang, S., Xiong, H., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9643. [CrossRef]

16. Agrawal, R.; Golshan, B.; Terzi, E. Grouping students in educational settings. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14), New York, NY, USA, 24–27 August 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 1017–1026. [CrossRef]

17. Bockmon, R.; Cooper, S.; Gratch, J.; Zhang, J.; Dorodchi, M. Can Students' Spatial Skills Predict Their Programming Abilities? In Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20), Trondheim, Norway, 15–19 June 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 446–451. [CrossRef]

18. Pardi, G.; von Hoyer, J.; Holtz, P.; Kammerer, Y. The Role of Cognitive Abilities and Time Spent on Texts and Videos in a Multimodal Searching as Learning Task. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20), Vancouver, BC, Canada, 14–18 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 378–382. [CrossRef]

19. Ghosh, S.; Rath, M.; Shah, C. Searching as Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks. In Proceedings of the CHIIR '18: Conference on Human Information Interaction and Retrieval, New Brunswick, NJ, USA, 11–15 March 2018; pp. 22–31. [CrossRef]

20. Demaree, D.; Jarodzka, H.; Brand-Gruwel, S.; Kammerer, Y. The Influence of Device Type on Querying Behavior and Learning Outcomes in a Searching as Learning Task with a Laptop or Smartphone. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20), Vancouver, BC, Canada, 14–18 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 373–377. [CrossRef]

21. Syed, R.; Collins-Thompson, K. Retrieval Algorithms Optimized for Human Learning. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17), Tokyo, Japan, 7–11 August 201; Association for Computing Machinery: New York, NY, USA, 2017; pp. 555–564. [CrossRef]

22. Syed, R.; Collins-Thompson, K.; Bennett, P.N.; Teng, M.; Williams, S.; Tay, W.W.; Iqbal, S. Improving Learning Outcomes with Gaze Tracking and Automatic Question Generation. In Proceedings of the Web Conference 2020 (WWW '20), Taipei, Taiwan, 20–24 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1693–1703. [CrossRef]

23. Roy, N.; Torre, M.V.; Gadiraju, U.; Maxwell, D.; Hauff, C. Note the Highlight: Incorporating Active Reading Tools in a Search as Learning Environment. In Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21), Canberra, Australia, 14–19 March 202; Association for Computing Machinery: New York, NY, USA, 2021; pp. 229–238. [CrossRef]

24. Mao, J.; Liu, Y.; Kando, N.; Zhang, M.; Ma, S. How Does Domain Expertise Affect Users' Search Interaction and Outcome in Exploratory Search? *ACM Trans. Inf. Syst.* **2018**, *36*, 42. [CrossRef]

25. El Zein, D.; Câmara, A.; Da Costa Pereira, C.; Tettamanzi, A. RULKNE: Representing User Knowledge State in Search-as-Learning with Named Entities. In Proceedings of the 2023 Conference on Human Information Interaction and Retrieval (CHIIR '23), Austin, TX, USA, 19–23 March 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 388–393. [CrossRef]

26. Liu, J.; Jung, Y.J. Interest Development, Knowledge Learning, and Interactive IR: Toward a State-based Approach to Search as Learning. In Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21), Canberra, Australia, 14–19 March 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 239–248. [CrossRef]

27. Collins-Thompson, K.; Rieh, S.Y.; Haynes, C.C.; Syed, R. Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies. In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16), Canberra, Australia, 14–19 March 2021; Association for Computing Machinery: New York, NY, USA, 2016; pp. 163–172. [CrossRef]
28. Zhu, H.; Tian, F.; Wu, K.; Shah, N.; Chen, Y.; Ni, Y.; Zhang, X.; Chao, K.-M.; Zheng, Q. A multi-constraint learning path recommendation algorithm based on knowledge map. *Knowl. Based Syst.* **2018**, *143*, 102–114. [CrossRef]
29. Zhou, X.; Chen, J.; Jin, Q. Discovery of Action Patterns in Task-Oriented Learning Processes. In *Advances in Web-Based Learning—ICWL 2013, Proceedings of the 12th International Conference, Kenting, Taiwan, 6–9 October 2013*; Wang, J.F., Lau, R., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8167. [CrossRef]
30. Mehrotra, R.; Yilmaz, E. Extracting Hierarchies of Search Tasks & Subtasks via a Bayesian Nonparametric Approach. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17), Tokyo, Japan, 7–11 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 285–294. [CrossRef]
31. Hastuti, R.P.; Suyanto, Y.; Sari, A.K. Q-Learning for Shift-Reduce Parsing in Indonesian Tree-LSTM-Based Text Generation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2022**, *21*, 64. [CrossRef]
32. Lin, J.-L.; Kuo, J.-C.; Chuang, H.-W. Improving Density Peak Clustering by Automatic Peak Selection and Single Linkage Clustering. *Symmetry* **2020**, *12*, 1168. [CrossRef]
33. Min, X.; Huang, Y.; Sheng, Y. Automatic Determination of Clustering Centers for "Clustering by Fast Search and Find of Density Peaks". *Math. Probl. Eng.* **2020**, *2020*, 4724150. [CrossRef]
34. Yang, Q.-F.; Gao, W.-Y.; Han, G.; Li, Z.-Y.; Tian, M.; Zhu, S.-H.; Deng, Y.-H. HCDC: A novel hierarchical clustering algorithm based on density-distance cores for data sets with varying density. *Inf. Syst.* **2023**, *114*, 102159. [CrossRef]
35. Ahmed, M.; Samee, M.; Mercer, R. Improving Tree-LSTM with Tree Attention. In Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 30 January–1 February 2019; pp. 247–254. [CrossRef]
36. Shido, Y.; Kobayashi, Y.; Yamamoto, A.; Miyamoto, A.; Matsumura, T. Automatic Source Code Summarization with Extended Tree-LSTM. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [CrossRef]
37. Yu, X.; Li, G.; Chai, C.; Tang, N. Reinforcement Learning with Tree-LSTM for Join Order Selection. In Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 20–24 April 2020; pp. 1297–1308. [CrossRef]
38. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef]
39. Su, C.; Huang, H.; Shi, S.; Jian, P.; Shi, X. Neural machine translation with Gumbel Tree-LSTM based encoder. *J. Vis. Commun. Image Represent.* **2020**, *1*, 102811. [CrossRef]
40. Lindemann, B.; Müller, T.; Vietz, H.; Jazdi, N.; Weyrich, M. A survey on long short-term memory networks for time series prediction. *Procedia CIRP* **2021**, *99*, 650–655. [CrossRef]
41. Tai, K.S.; Socher, R.; Manning, C.D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; Volume 1, pp. 1556–1566. [CrossRef]
42. Blundell, C.; Teh, Y.W. Bayesian hierarchical community discovery. In Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS '13); Curran Associates Inc.: Red Hook, NY, USA, 2013; Volume 1, pp. 1601–1609.
43. Parmar, D.; Wu, T.; Blackhurst, J. MMR: An algorithm for clustering categorical data using Rough Set Theory. *Data Knowl. Eng.* **2007**, *63*, 879–893. [CrossRef]
44. Jin, W. Graph Mining with Graph Neural Networks. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21); Association for Computing Machinery: New York, NY, USA, 2021; pp. 1119–1120. [CrossRef]
45. Ma, J.; Gao, W.; Joty, S.; Wong, K.-F. An Attention-based Rumor Detection Model with Tree-structured Recursive Neural Networks. *ACM Trans. Intell. Syst. Technol.* **2020**, *11*, 42. [CrossRef]