Education Data Taking out to Carry Program Learning via Analytic Data

R. ARUNKUMAR
Dr. N.SATHYABALAJI M.E., Ph.D

Department of Computer Science and Engineering, JKK Munirajah College of Technology, Erode, Tamilnadu 638 506, India

Abstract:
The framework includes the following sequence data collection, log and score collection, and preprocessing clustering algorithms for problem-solving, for clustering data in Euclidean space, extracting statistical features from each cluster, and applying A frequent pattern growth algorithm for each cluster to determine data patterns and association rules. A set of recommendations based on extracted features, data patterns and rules. Tune other parameters to get the best results for the clustering and association rule mining algorithms.

At this time, in addition to classroom learning, the system also enhances programming learning and hands-on opportunities unsupervised algorithms.

The classification generates a large archive of analytical data result codes, logs, and score that can be valuable raw material for research in programming learning. We propose an educational data mining structure that supports programming learning using

Algorithms and data structures are used in programming courses to approximate the data that students use to solve real-world problems. We also experimentally use synthetic data to demonstrate the performance of our algorithm. Investigational results demonstrate that the proposed framework successfully extracts useful features, patterns, and rules from analytical data. Additionally, these extracted features, patterns, and rules highlight weaknesses and areas for improvement in programming learning.

Computer programming is of broad interest in the development of information and communication technologies in real life. Meet the growing demand for highly skilled programmers in critical challenges.

Index: *Education, Carry Program, Analytic Data, Data collection*
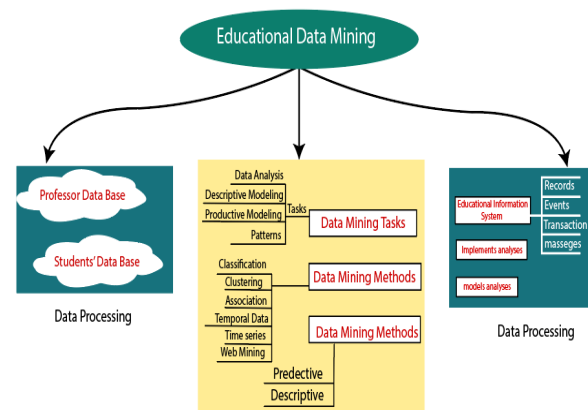
I.   Introduction:

Today's information and communication technology (ICT) industry demands for highly skilled programmers for further development. The conventional computer programming learning environment is insufficient to prepare highly skilled programmers due to the limited number of exercise classes, limited practice opportunities, and lack of individual tutoring. In addition, most educational institutions, such as schools, colleges, and universities are struggling to build more educational facilities to increase academic activity (e.g., additional exercise classes, practice, and individual tutoring) due to logistical and organizational constraints . The growing number of people in classrooms in educational institutions.

The large number of students per class, and some lectures are conducted with more than a thousand participants in the massive open online courses which complicate the individual tutoring process. Moreover, the growing ratio between students and educator raises the question of how to

supply personality support to students to improve their analytical skills. Especially, when learning computer programming, students need a lot of practice and personality tutoring to improve their training knowledge and skills.
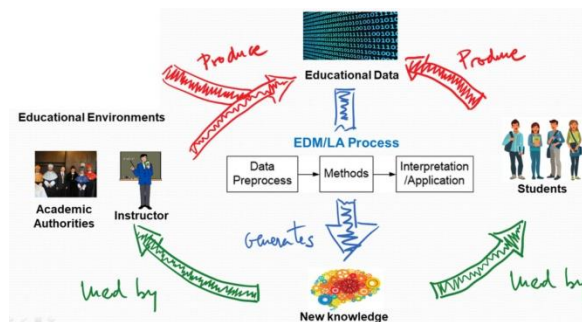
Computer programming is one of the fundamental courses in ICT discipline. Programming practice and competition can play an important role in acquiring good programming skills.

They proposed an evaluation approach that compares the strength of clustering algorithms on a given dataset and automatically recommends a suitable algorithm for EDM. Unlike existing models, the proposed framework for EDM leverages real-world problem solving data collected from an OJ system to find data patterns, rules, and features. Finally, some useful recommendations for learners and educators are generated based on the identified features, rules, and data patterns.

To the best of our information, most of the on hand systems use a learning-by-doing formula to support student training skills. The huge volume of data stored in these systems helps researchers find shortcomings in students training and identify areas for enhancement. A lot of research is being conducted using these rich property to identify and solve various programming-related problems.
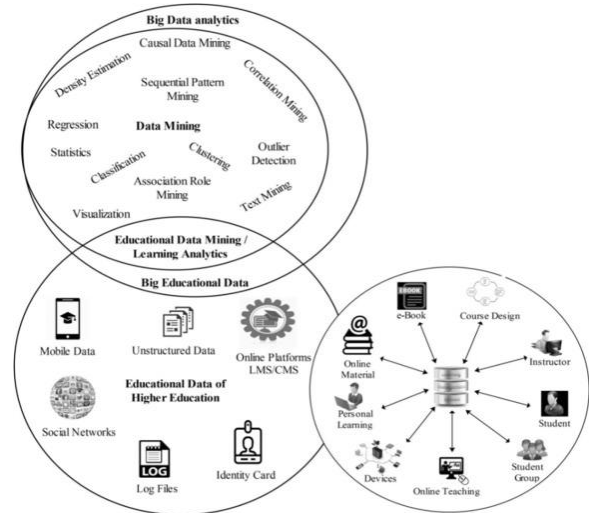
This paper makes the following contributions:
• We propose a framework for EDM and data analysis using unsupervised machine learning algorithms. To demonstrate the effectiveness of the proposed framework, experiments are conducted on real-world problem-solving dataset.
• Various features, patterns, and rules are extracted from problem-solving data to support programming learning. Useful recommendations are generated for students and educators on the basis of the extracted features, rules, and data patterns.
• The extracted knowledge (features, rules and patterns) shows the weaknesses and the scope of possible improvements in programming learning.
• The proposed EDM framework can be integrated into e-learning platforms and OJ systems.

To the best of our knowledge, there is no single clustering technique that can handle all types of data including text, numbers, images, and videos. Conducted a comprehensive survey on clustering techniques and discussed their advantages, disadvantages, evaluations, and their complexity. Clustering techniques can be classified into different groups (hierarchy, fuzzy theory, distribution, density, graph theory, grid, fractal theory, and model) based on their working procedures.

Comparing Apriori and FP-growth, Apriori requires repeated scanning of the database to form a candidate item set, whereas the FP-growth algorithm is very fast because it only needs to scan the database twice to complete the process. Many variations and improvements of both algorithms have been introduced in different studies, including hashing technique, sampling approach, dynamic counting depth-first mining, h-mine, and tree structures

These systems are hosted by big technology giants to receive voice or text data and produce the relevant output. Despite the many advantages of SPA systems for quick answers, they are difficult to employ for the purpose of programming learning and evaluation.

Clustering techniques are widely used in data analysis and play an important role in the field of data mining. With the diversification of data, many variations of clustering techniques have been developed simultaneously to analyze different types of data. Each clustering technique has its advantages and disadvantages for clustering data. The usability and applicability of clustering techniques in the context of the EDM has been described in a study.

In the proposed approach, the FP-growth algorithm is leveraged for mining patterns and association rules. The FP-growth algorithm is faster than other algorithms and has the advantage of generating candidate item sets after only two repetitions (scans) of the database

Educational data mining

The rapid proliferation of e-learning platforms and their use in educational institutions has resulted in a huge amount of data archives. Educational data is evolving and the diversity of these data varies from one e-learning platform to another.
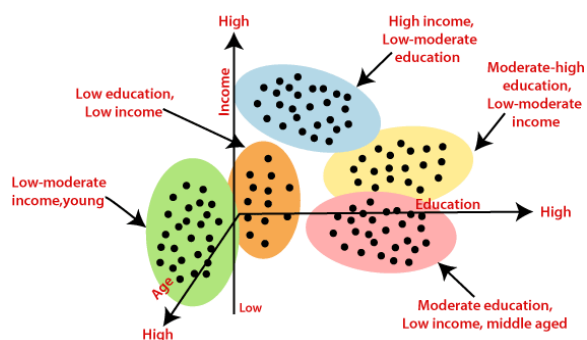
Efficiently handling this massive and diverse educational data is a non-trivial and challenging task. Educational data mining (EDM) technique has emerged to address this problem. Due to the variety, volume, nature, and structure of learning data, conformist data mining algorithms cannot be applied straight.

The experimental results of EDM using problem-solving data can be integrated to visualize different LA for programming platforms such as the OJ system. In addition, fuzzy estimation and polynomial approximation methods can be handy

to dynamically select the optimal minSup values based on the dataset. Appropriate values could help to generate the actual number of frequent elements and association rules from the dataset.

The MK-means clustering algorithm is used to process the multidimensional education data, select the optimal initial centric, and handle outliers. These features make this clustering algorithm more effective than other partition-based clustering algorithms.

Data analysis, including data patterns, feature extraction, and association rules, is closely related to data clustering, and improper clustering can lead to poor results. Before clustering our multidimensional education data, we analyzed the suitability of different clustering methods and algorithms for our clustering objectives. Each algorithm has its own suitability for a particular application, so it is very difficult to find the best clustering algorithm. However, we looked at the hierarchical clustering method, which has high computational and memory requirements compared to the partitioned clustering methods.

One of the main objectives of this research is to understand what difficulties students have in solving programming problems, identify the main influencing factors in their programming learning process, and determine what strategies, methods, or technologies can be used in teaching and learning to improve students' programming skills

Programming learning

Computer programming has attracted a lot of attention in the development of information and communication technologies in the real world. Meeting the growing demand for highly skilled programmers in the ICT industry is one of the major challenges. In this point, online judge (OJ) systems enhance training learning and practice opportunity in addition to classroom-based learning. Therefore, OJ systems have created a large number of analytical data solution codes, logs, and scores archives that can be valuable raw materials for training learning research.

Pattern and Association

an unsupervised algorithm and was rst introduced inresearch. There are diverse applications of the ARM technique in various elds such as pattern mining, education, social, medical, census, market-basket, and big data analysis.

ARM is an efcient technique for obtaining frequent items from large datasets. Among the many types of ARM algorithms, Apriori and FP-growth algorithms are the most widely used. Comparing Apriori and FP-growth,

Apriori requires repeated scanning of the database to form a candidate itemset, whereas the FP-growth algorithm is very fast because it only needs to scan the database twice to complete the process. Many variations and improvements of both algorithms have been introduced in different studies, including hashing technique, sampling approach, dynamic counting, depth-rst mining, h-mine, and tree structures

Educational Data Mining

e-learning platforms have become more popular for a variety of reasons and demands, including teacher shortage, unbalanced student-teacher ratio, logistical and infrastructure constraints, high cost of technical and professional courses, dissemination of education to a large number of people, time saving and easy access to many courses. As the use of e-learning systems increases, different types of data are being generated regularly.

Some data are structured whereas some are unstructured. Therefore, it is very difcult to retrieve useful information from this huge amount of mixed data archives using traditional

Problem-solving data

In the performance of clustering algorithms is automatically compared and the most efficient clustering algorithm is recommended based on the dataset. Statistical data analysis is performed on programming logs and test scores to investigate the hidden information of programmers in the study.
Furthermore, EDM can be employed to find useful knowledge and information from problem-solving data to support programming learning.

Although, extracting hidden information and discovering data patterns and rules from the problem-solving data is achallenging task. In this paper, we have exploited unsupervised algorithms to extract useful information and find data patterns from real-world problem-solving data collected from system.

Clustering

We have exploited unsupervised algorithms to extract useful information and find data patterns from real-world problem-solving data collected from an OJ system. Upon completion of data preprocessing task, a clustering algorithm is applied to the processed data to cluster it and then statistical features are extracted from each cluster. Next, the frequent pattern (FP)-growth algorithm is applied to explore data patterns and association rules from each cluster.

It is important to find more meaningful features among the many statistical hypothesized features that can be useful for improving programming learning.

The following statistical features can be generated in each cluster to explore the hidden information from the problem-solving data.

We have highlighted some important features: tendency to submit each assignment on a daily basis, average, standard deviation, and variance of scores in different tests

Algorithm:

In this paper, we propose an educational data mining framework to support programming learning using unsupervised algorithms. The

framework includes the following sequence of steps:

(i) problem-solving data collection (logs and scores are collected from the OJ) and preprocessing;

(ii) MK-means clustering algorithm is used for data clustering in Euclidean space;

(iii) Statistical features are extracted from each cluster;

(iv) Frequent pattern (FP)-growth algorithm is applied to each cluster to mine data patterns and association rules;

(v) A set of suggestions are provided on the basis of the extracted features, data patterns, and rules.

The experiment, the FP-growth algorithm is used to discover the frequent data patterns in each cluster. First, the frequency of different attributes in each cluster is calculated,

Then, the ranking of the attributes based on frequency is also enumerated. Next, we compute the frequent data patterns for each cluster by varying the minimum support (minSup) value.

K-means clustering

He clustering process, ODM is another effective module (Algorithm 2) that can detect and remove the most irrelevant data points from the dataset. In this way, the probability of selecting irrelevant data points as optimal initial centres is reduced. Figure 2 shows an example of two-dimensional data distribution in Euclidean space, where the distance of the data points from the origin can be calculated he key phases is data clustering. Data analysis, including data patterns, feature extraction, and association rules, is closely related to data clustering, and improper clustering can lead to poor results.

They proposed an evaluation approach that compares the strength of clustering algorithms on a given dataset and automatically recommends a suitable algorithm for EDM. Unlike existing models, the proposed framework for EDM leverages real-world problem- solving data collected from an OJ system to data patterns, rules, and features. Finally, some useful recommendations for learners and educators are generated based on the indented features, rules, and data patterns.

K-means clustering was used to evaluate students' activities in an e-learning system and indented students' interests. It also indented the correlation between the activity in the e-learning system and academic performance.

Before clustering our multidimensional education data, we analyzed the suitability of different clustering methods and algorithms for our clustering objectives. Each algorithm has its own suitability for a particular application, so it is very difficult to find the best clustering algorithm. However, we looked at the hierarchical clustering method, which has high computational and memory requirements compared to the partitioned clustering methods. For large data sets,a hierarchical algorithm can be very expensive. The main differences between hierarchical and partition methods are computation time, prior assumptions, data sets, and cluster.

Clustering techniques are widely used in data analysis and play an important role in the field of data mining. With the diversification of data, many variations of clustering techniques have been developed simultaneously to analyze different types of data. Each clustering technique has its advantages and disadvantages for clustering data. The usability and applicability of clustering techniques in the context.

Rule-Based Recommendation Systems

The volume and variety of content on e-learning platforms are increasing at an unprecedented rate, and at the same time the opportunities for research using the resources of e-learning platforms are also increasing. Recommending relevant and appropriate content to users (e.g., students, instructors, and teachers) is a challenging and tough task for any e-learning platform. Perusal proposed a novel personalized RS to provide appropriate supportive content to users. In their approach, FP-growth algorithm is applied to generate frequent items patterns, and fuzzy logic is used to partition the content into three levels. Recently, some RSs have been using a mixed approach of content-based ltering and collaborative ltering to achieve high-quality results in contexts. In addition, most RSs are built with a collaborative, knowledge-based, content-based, and hybrid approaches. Conventional e-learning platforms are insufficient to assess exercise-based content such as programming solution codes automatically, instead they can assess exercise-based contents semi-automatically. Thus, usual RS in e-learning platforms have limited suitability for programming and exercise-based education.

we have exploited unsupervised algorithms to extract useful information and data patterns from real-world problem-solving data collected from an OJ system. Upon completion of data pre-processing task, a clustering algorithm is applied to the processed data to cluster it and then statistical features are extracted from each cluster. Next, the frequent pattern (FP)-growth algorithm is

applied to explore data patterns and association rules from each cluster.

Finally, recommendations are provided for students and educators based on the extracted features, patterns, and rules. In particular, we have focused on ending effective data patterns from problem-solving data that can be useful to support programming learning. This paper makes the following contributions.

E-learning platforms have become more popular for a variety of reasons and demands, including teacher shortage, unbalanced student-teacher ratio, logistical and infrastructure constraints, high cost of technical and professional courses, dissemination of education to a large number of people, time saving and easy access to many courses. As the use of e-learning systems increases, different types of data are being generated regularly. Some data are structured whereas some are unstructured. Therefore, it is very difficult to retrieve useful information from this huge amount of mixed data archives using traditional

II. Conclusion:

We proposed an EDM framework for data clustering, patterns, and rules mining using real-world problemsolving data. A mathematical model for data preprocessing, MK-means, and FP-growth algorithms were used to conduct this study. For programming education, OJ systems have been adopted by many institutions as academic tools. As a result, a huge number of programming-related resources source codes, logs, scores, activities, are regularly accumulated in OJ systems. In this study, a large amount of real-world problem-solving data collected from the AOJ system was used in the experiments.

Problem-solving data preprocessing is one of the main tasks to achieve accurate EDM results. Therefore, a mathematical model for problem-solving data preprocessing is developed. Then, the processed data are clustered using Elbow and MK-means algorithms. Various statistical features, data patterns and rules are extracted from each cluster based on different threshold values (K, minConf , minSup).

Future Work:

The experimental results of EDM using problem-solving data can be integrated to visualize different LA for programming platforms such as the OJ system. In addition, fuzzy estimation and polynomial approximation methods can be handy to dynamically select the optimal minSup values based on the dataset. Appropriate minSup values could help to generate the actual number of frequent elements and association rules from the dataset.

Reference:

[1]. R. Rietsche, K. Duss, J. M. Persch, and M. Soellner, ''Design and evaluation of an IT-based formative feedback tool to foster student performance,'' in Proc. 39th Int. Conf. Inf. Syst. (ICIS), San Francisco, CA, USA, 2018, pp. 1–17.

[2]. S. Oeste, K. Lehmann, A. Janson, M. Sollner, and J. M. Leimeister, ''Redesigning university large scale lectures: How to activate the learner,'' in Proc. 75th Acad. Manage. Annu. Meeting, Vancouver, BC, Canada, 2015, pp. 1–5.

[3]. J. A. Kulik and J. D. Fletcher, ''Effectiveness of intelligent tutoring systems,'' Rev. Educ. Res., vol. 86, 42–78, May 2016, doi: 10.3102/0034654315581420.

[4]. W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, ''Intelligent tutoring systems and learning outcomes: A meta-analysis,'' J. Educ. Psychol., vol. 106, 901–918, May 2014, doi: 10.1037/a0037123.

[5]. J. C. Nesbit, O. O. Adesope, Q. Liu, and W. Ma, ''How effective are intelligent tutoring systems in computer science education?'' in Proc. IEEE 14th Int. Conf. Adv. Learn. Technol., May 2014, pp. 99–103.

[6].   K. Vanlehn, ''The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,'' Educ. Psychol., vol. 46, pp. 197–221, Oct. 2011, doi: 10.1080/00461520.2011.611369.

[7].   A. Janson, M. Sollner, and J. M. Leimeister, ''Ladders for learning: Is scaffolding the key to teaching problem-solving in technology mediated learning contexts,'' Acad. Manage. Learn. Educ., vol. 19, no. 4, 439–468, 2020.

[8].   M. C. Kim and M. J. Hannafin, ''Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice,'' Comput. Educ., vol. 56, 403–417, Feb. 2011.

[9].   T. A. Brush and J. W. Saye, ''A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment,'' J. Interact. Learn., vol. 1, pp. 1–12, Oct. 2002.

[10]. R. Knote, A. Janson, and J. M. Leimeister, ''Value co-creation in smart services: A functional affordances perspective on smart personal assistants,'' J. Assoc. Inf. Syst., vol. 4, pp. 418–458, Mar. 2020.