Design of an Online Data Structure Practice System Based on K-means Algorithm

Liu Yan, Long Yanbin*
Liaoning University of Science and Technology, Anshan, China
**Corresponding Author*

Abstract: This paper focuses on an online data structure practice system based on the K-means algorithm, and explores its application value in computer science education. Through a detailed explanation of the system design and implementation process, including the specific application of the K-means algorithm in data clustering, system architecture design, and functional module division, it demonstrates how the system effectively assists students in mastering data structure knowledge. Simultaneously, combined with practical application cases, it analyzes the system's significant effectiveness in improving student learning outcomes and teacher teaching efficiency, providing new ideas and methods for online data structure education.

Keywords: K-means Algorithm; Data Structure; Online Practice System; Educational Application

1. Introduction

With the rapid development of information technology, online education has gradually become an important part of the education field. As a core course in computer science, improving the teaching quality of data structures is crucial for cultivating students' programming and logical thinking abilities. However, traditional data structure teaching methods are often limited by time and space, making it difficult to meet the needs of students' personalized learning. Therefore, developing an online data structure practice system based on the K-means algorithm has significant practical implications [1] .

The K-means algorithm, as a classic unsupervised learning clustering algorithm, has the advantages of simplicity, efficiency, and ease of implementation, and can effectively classify and cluster data. Applying the K-means algorithm to an online data structure practice system can realize intelligent analysis of student learning data, provide students with personalized learning suggestions and practice questions, thereby improving students' learning outcomes [2] .

2. Overview of K-means Algorithm

2.1 Algorithm Principle

The K-means algorithm is a distance-based clustering algorithm. Its basic idea is to divide the dataset into k clusters, such that each data point belongs to the cluster corresponding to the nearest cluster center. The algorithm iteratively optimizes the position of the cluster centers until the convergence condition [3-4] is met. The specific steps are as follows:

Randomly select k data points as initial cluster centers.

Calculate the distance from each data point to the center of each cluster and assign it to the nearest cluster.

Recalculate the center point of each cluster, i.e., the mean of all data points within the cluster.

Repeat steps 2 and 3 until the cluster center no longer changes or the preset number of iterations is reached.

2.2 Algorithm Advantages and Disadvantages

Advantages: The K-means algorithm has advantages such as high computational efficiency, simple implementation, and applicability to large-scale datasets. It can quickly divide the dataset into multiple clusters, providing convenience for subsequent data analysis and processing [5] .

Disadvantages: The K-means algorithm requires pre-specifying the number of clusters k, and is sensitive to the selection of initial cluster centers, easily getting trapped in local optima. In addition, the algorithm is sensitive to noise and outliers, which may affect the accuracy of the clustering results.

3. Design of an Online Data Structure Practice System Based on K-means Algorithm

3.1 System Requirements Analysis

The online data structure practice system based on K-means algorithm is mainly aimed at computer science students and teachers, aiming to provide a convenient and efficient learning and teaching platform. The system should have the following functions:

Student Functions: Students can log in to the system for online practice, view practice questions and answer explanations, submit assignments, and view grades. Simultaneously, the system should provide personalized learning suggestions and practice questions [6] based on students' learning data.

Teacher Functions: Teachers can manage course information, publish practice questions and assignments, and view students' learning progress and grade statistics. In addition, teachers can use the system's data analysis functions to understand students' learning difficulties and weaknesses, so as to adjust teaching strategies [7-8] .

3.2 System Architecture Design

The system adopts a B/S (Browser/Server) architecture, which has the advantages of strong distribution, convenient maintenance, simple development and strong sharing, and low total cost of ownership [9] . The entire system is mainly composed of three closely cooperating parts: client, server and database. Its architecture diagram 1 is as follows:

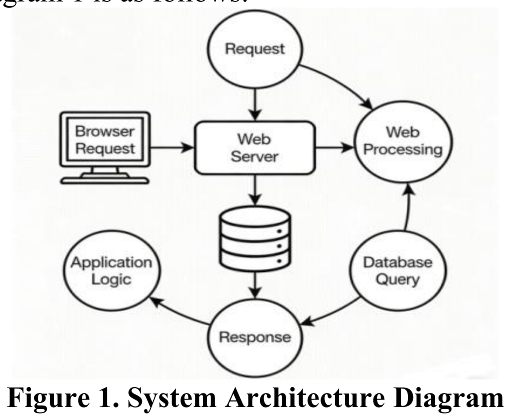

Figure 1. System Architecture Diagram

In the system architecture diagram 1, the top layer is the client layer, which uses a browser as the carrier to provide users with a user-friendly and intuitive interactive interface. Both students and teachers can access the system through a browser and perform various operations. The client interface design focuses on user experience, adopting a simple and clear layout and operation process to facilitate users to get started quickly. For example, after logging into the system, students can clearly see the course list, practice entry, etc.; after logging into the system, teachers can easily manage courses, publish assignments, etc. [10] .

The middle layer is the server-side, which is the core processing part of the system. The server is responsible for receiving various requests from the client, such as practice answers submitted by students, course information published by teachers, etc. After receiving the request, the server-side will perform business logic processing, such as verifying user identity, checking whether the submitted homework meets the requirements, recommending relevant questions based on students' answer situation, etc. After processing the business logic, the server-side will interact with the database to read or store the corresponding data. For example, when students register, the server will store the information filled in by the students in the database; When teachers query student grades, the server will read relevant data from the database and return it to the teacher.

At the bottom is the database layer, which is used to store various data of the system. A database is like a huge warehouse, storing important data such as user information, course information, practice questions, student assignments, and learning records. These data are organized and managed in a structured manner, facilitating quick queries and updates on the server side. For example, the user information table records each user's account, password, name, role (student or teacher), etc; The course information table includes the course name, course description, teaching outline, and instructor; The practice question table stores attributes such as the content, difficulty, type, and score of the questions; The student homework sheet records the content, submission time, grades, etc. of the homework submitted by students; The learning record sheet tracks students' learning process, such as answer time, accuracy, etc.

3.3 Functional Module Design

User Management Module

This module is mainly responsible for the full lifecycle management of users, including user registration, login, information modification, and other operations. During user registration, the system provides a detailed registration form, requiring users to fill in necessary personal information such as username, password, email, and mobile phone number, and performs format verification to ensure that the information entered by the user conforms to the specifications. At the same time, to ensure the security of user information, the system uses encryption technology to encrypt and store sensitive information such as user passwords. When a user logs in, the system verifies whether the user's entered account and password match; only users who pass the verification can successfully log in to the system. After successfully logging in, if a user needs to modify personal information, such as changing their password or updating contact information, they can perform convenient operations within the system. The entire user management module ensures the security and accuracy of user information through strict access control and data encryption mechanisms. Its flowchart 2 is as follows:

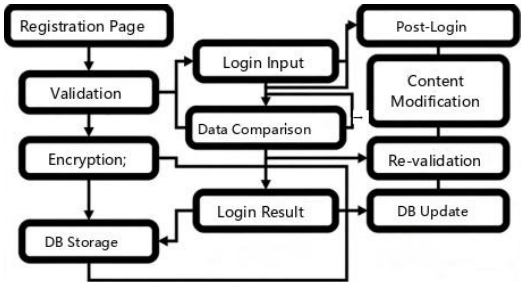

Figure 2. User Management Module Flowchart

As can be seen from flowchart 2, the user first enters the registration page and fills in information. The system performs format verification; if it fails, the user is prompted to re-enter the information; if it passes, the encrypted information is stored in the database. When a user logs in, they enter their username and password. The system compares this information with the encrypted information stored in the database. If they match, login is allowed; otherwise, a login failure message is displayed. When a user modifies information, they must log in first, then submit the changes. The system then verifies and encrypts the information again before storing it.

Course Management Module

The course management module provides comprehensive course management functions for teachers, who can create, edit, and delete course information. When creating a course, teachers need to fill in detailed information such as course name, course description, and teaching outline, which will help students better understand the course content and objectives. For example, the course name should be concise and clear, accurately reflecting the core content of the course; The course description can provide a detailed introduction to the target audience and learning outcomes of the course; The teaching syllabus can list the main knowledge points and learning progress of the course in the form of chapters or themes. After the teacher creates the course, students can view the course information on the client and choose courses that interest them based on their interests and needs for learning. If teachers need to modify the created courses, such as updating teaching content, adjusting teaching progress, etc., they can perform editing operations in the system. Meanwhile, if a course is no longer offered for various reasons, the teacher can also delete it.

In the interface, the top is the course list, displaying all created course names, introductions, and other information; There is a "Create Course" button below, click it to pop up a form for filling in course details; For each course in the list, there are "Edit" and "Delete" buttons for teachers to perform corresponding operations.

Practice Question Management Module

The exercise question management module endows teachers with the ability to flexibly manage exercise questions. Teachers can publish new practice questions, and when publishing, they need to set the difficulty, type, score, and other attributes of the questions. The system supports multiple question types, such as multiple-choice questions, fill in the blank questions, programming questions, etc., to meet the needs of different knowledge points and teaching objectives. For example, for some basic conceptual knowledge points, multiple choice questions or fill in the blank questions can be set up; For knowledge points that require students' hands-on practice and programming skills, programming questions can be set up. After teachers publish questions, they can also edit and delete them based on teaching feedback. If a question is found to have errors or needs to be updated, the teacher can promptly edit and modify it; If a question is no longer applicable, it can also be deleted. The data structure example of the exercise question management module is as follows (presented in table form):

Table 1. Data Structure Example Table for the Exercise Question Management Module

Question ID	Question Content	Difficulty	Type	Score	Related Knowledge Point
1.	Which of the following options is a linear data structure? A. Tree B. Graph C. Stack D. Heap	Easy	Multiple Choice Question	5	Linear Structure
2	Please fill in the basic operations of a stack in data structures. The Last-In-First-Out (LIFO) operation is _.	Medium	Fill in the Blank Question	3	Stack Operations
3	Write a program to store integers in a linked list and sort them.	Difficult	Programming Question	15	Linked List Sorting Algorithm

Online Practice Module

The online practice module is an important part of students' knowledge consolidation and self-testing. Students can practice online, and the system will record students' answers in real time, including answer time and answer content. After students submit their answers, the system immediately provides answer explanations and feedback to help them understand their performance and identify the reasons for errors. For example, for multiple-choice questions, the system displays the correct answer and explains the meaning of each option; for programming questions, the system provides the code execution results and error messages to guide students in debugging and modification. Simultaneously, the system recommends relevant practice questions based on students' performance. If a student has a low accuracy rate on a particular knowledge point, the system recommends more questions on that knowledge point for practice to help strengthen their weaknesses. An example of the online practice module's answer interface is shown in Figure 3:

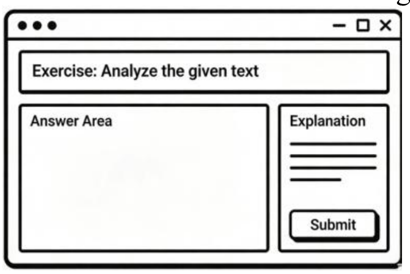

Figure 3. Answer Interface of the Online Practice Module

In the interface, the question content is displayed at the top, the answer area is in the middle (depending on the question type, it may be a selection box, fill-in-the-blank box, or code editing box), and there is a "Submit" button at the bottom. After submission, answer explanations and feedback information are displayed.

Homework Management Module

The homework management module realizes the automated process of assigning homework by teachers and submitting homework by students. Teachers can assign homework, set deadlines and submission requirements for homework, such as format and word limit. After the teacher assigns homework, students can view the homework details in the system and submit the homework according to the requirements. For the objective question section, the system will automatically correct and quickly provide grades, saving teachers' time and energy; For subjective questions, teachers can view students' answers in the system, manually correct them, and provide detailed grades and comments to provide timely feedback on students' learning progress.

In the interface, a list displays the names, deadlines, submission status (submitted/not submitted), and other information of all assignments. Students can click on the assignment name to view detailed requirements and perform submission operations; Teachers can click on homework to grade and view statistical information.

Data Analysis Module

The data analysis module is the intelligent core of the system. It uses the K-means algorithm to perform cluster analysis on students' learning data. Students' learning data includes multi-dimensional information such as answer accuracy, answer time, number of practice sessions, and homework scores. Through the K-means algorithm, the system can divide students into different learning groups and explore students' learning patterns and characteristics. For example, students can be divided into three clusters: strong, average, and weak learning abilities, or into diligent, average, and lazy learning types based on their learning habits. Teachers can understand each student's learning situation based on the analysis results and provide personalized learning suggestions for different learning groups or individual students. For example, for students with strong learning abilities, more challenging learning resources and extension questions can be recommended; for students with weak learning abilities, more basic tutoring and targeted practice can be provided.

For example, design a bar chart where the horizontal axis represents different learning groups (e.g., high, medium, and low learning abilities) and the vertical axis represents the number of students. The height of the bars visually displays the distribution of students in each group. Alternatively, design a line chart where the horizontal axis is time (e.g., week number) and the vertical axis is the average student score, showing the trend of student scores over time and helping teachers understand teaching effectiveness and student learning progress.

4. Implementation of an Online Practice System Based on the K-means Algorithm

4.1 Development Environment and Technology Selection

The system is developed using Java, with the server-side built using the Spring Boot framework and data stored in a MySQL database. The front-end uses HTML, CSS, and JavaScript to implement the user interface and uses Ajax technology for asynchronous communication with the server.

The following table details the configuration information of the system development environment:

Project Details
Development Language Java 11
Development Framework Spring Boot 2.5.0
Database MySQL 8.0
Front-end Technology HTML5 、 CSS3 、 JavaScript
Development Tools IntelliJ IDEA 2021.3
Server Tomcat 9.0
Build Tools Maven 3.8.1

4.2 Application of K-means Algorithm in the System

In the system, the K-means algorithm is mainly used for cluster analysis of student learning data. The specific steps are as follows:

Data Collection: Collect student learning data, including answer accuracy, answer time, number of practice sessions, etc. The table below shows some examples of collected student learning data:

Table 2. Student Learning Data

Student ID	Answer accuracy	Answering time (seconds)	Number of practice sessions
1001	0.85	120	15
1002	0.72	180	10
1003	0.92	90	20
1004	0.65	210	8
1005	0.88	110	18

Data preprocessing: The collected data is cleaned and standardized to remove noise and outliers, and the data is converted into a format suitable for K-means algorithm processing. For example, the correct answer rate is normalized to be between 0 and 1.

Determining the number of clusters k: The Elbow Method is used to determine the optimal number of clusters k. The Elbow Method observes the changes in the sum of squared errors (SSE) under different k values and selects the point where the rate of decrease in SSE drops sharply as the k value. Figure 4 below shows the changes in SSE under different k values:

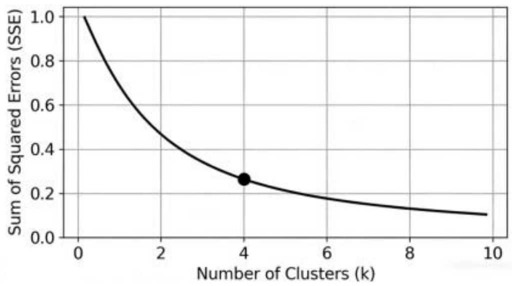

Figure 4. Shows the Changes in SSE under Different K Values

Initialize cluster centers: Randomly select k data points as initial cluster centers.

Iterative Clustering: According to the iterative process of the K-means algorithm, the cluster centers and data point assignments are continuously updated until the convergence condition is met. Its flowchart is shown in Figure 5 below:

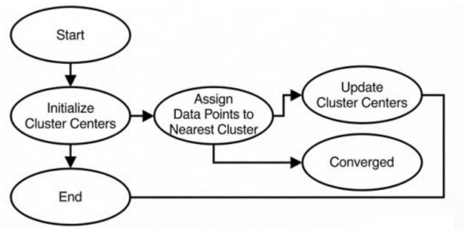

Figure 5. Iterative Process Flow of K-means Algorithm

Result Analysis: Analyzing the clustering results helps understand the learning characteristics and patterns of students in different clusters, providing personalized learning suggestions for students. For example, for students with high answer accuracy, short answering time, and many practice sessions, more challenging learning content can be recommended; for students with low answer accuracy, long answering time, and few practice sessions, more basic tutoring and practice can be provided.

4.3 Key Code Implementation

The following is some key code of K-means algorithm implemented in Java: (omitted)

The above code implements the core functions of the K-means algorithm, including calculating Euclidean distance, initializing cluster centers, assigning data points to clusters, updating cluster centers, and determining whether cluster centers are equal. By running this code, cluster analysis can be performed on given data to obtain the final cluster centers. In practical applications, student learning data can be used as input, and personalized learning suggestions can be provided to students based on the clustering results.

5. Application Case Analysis

5.1 System Application

After introducing an online data structure practice system based on the K-means algorithm, a university's computer science major conducted online practice for data structure courses on the system. The system recorded students' answers, learning time, and other data, and used the K-means algorithm to perform cluster analysis on this data.

5.2 Clustering Result Analysis

Through K-means algorithm clustering, students are divided into different learning groups. For example, they can be divided into three clusters: strong learning ability, medium learning ability, and weak learning ability. For students with strong learning abilities, the system recommends more challenging practice problems to further improve their programming skills; for students with average learning abilities, the system provides targeted tutoring and practice to help them consolidate their knowledge; for students with weak learning abilities, the system provides more basic exercises and detailed answer explanations to help them gradually master data structure knowledge.

5.3 Teaching Effectiveness Evaluation

After one semester of teaching practice, it was found that after using this system, students' grades in the data structure course improved significantly. At the same time, students' interest and enthusiasm for learning the data structure course also increased significantly. Teachers can better understand students' learning progress through the system's data analysis functions, adjust teaching strategies in a timely manner, and improve teaching efficiency.

6. Conclusion

6.1 Conclusion

This paper studies an online data structure practice system based on the K-means algorithm. Through the design and implementation of the system, the K-means algorithm is applied to the clustering analysis of student learning data, providing students with personalized learning suggestions and practice problems. Practical application cases show that the system can effectively improve students' learning outcomes and teachers' teaching efficiency, and has high application value.

6.2 Outlook

Future research can further optimize the K-means algorithm to improve the accuracy and efficiency of clustering. At the same time, it can be combined with other machine learning algorithms, such as decision trees and neural networks, to achieve more intelligent learning analysis and recommendation. In addition, the system's functions can be expanded, such as adding online communication and virtual laboratory modules, to provide students with more comprehensive learning support.

Acknowledgements

Supported by Liaoning University of Science and Technology 2025 Undergraduate Innovation and Entrepreneurship Project

References

[1] Zou Hui; Wang Jiandong; Cui Zhongqiang. Application research of master-slave multi-chain in college grade management. Fujian Computer, 2025(02).45-46

[2] Zhu Caidie; Lao Dongqing; Wu Gang. Research and practice on teaching reform of “Data Structure” course. Western Quality Education, 2025(19).3-5
[3] Mu Yingxue; Liang Jianming; Wen Xiumei. Exploration and practice of project-driven teaching model in “Data Structure” course. Science and Technology Wind, 2025(25).99-102
[4] Sun Daiju; Li Liangliang; Li Junping; Zhang Zhansheng; Guo Huanlong; Wang Xiaofeng. A method for structural conversion of geotechnical data [P]. Shandong Zhengyuan Construction Engineering Co., Ltd. 2025.56-58
[5] Li Xiang; Huang Jia; Yao Qiming. Research on identification and evolution analysis of standard digital hotspot technology. Standard Science, 2025(01)34-36
[6] Qu Qianyu. Carousel content display method, device, equipment and storage medium [P]. Tencent Technology (Shenzhen) Co., Ltd. 2025.77-79
[7] Qiu Xin. Data processing method, device, equipment and readable storage medium [P]. Tencent Technology (Shenzhen) Co., Ltd. 2025.81-83
[8] Song Haiyun; Wang Xiao; Zhang Pei; Lao Tian; Wang Xin; Yue Xinghua; Zheng Xiaodong; Liu Lei. A Sensitive Data De-identification Method [P]. China Railway Signal & Communication Corporation Limited. 2025. 34-38
[9] H. Rudolf; C.A. Winkler; I. Windler; M. Hartmann; R. Parker. Processing primary products into secondary products [P]. BASF Europe. 2025. 56-58
[10] H. Rudolf; M. Hartmann; I. Windler; R. Parker; C.A. Winkler. Processing primary products into secondary products [P]. BASF Europe. 2025. 34-38

