Research Article

A Study on Integrating Multimodal English and American Literature Resources Using Data Mining

Lanlan Cai1 and Zhiqiang Zhuang2

1School of Foreign Languages, Quanzhou Normal University, Quanzhou, Fujian 362000, China
2School of Foreign Languages, Minnan Normal University, Zhangzhou, Fujian 363000, China

Correspondence should be addressed to Zhiqiang Zhuang; zzq0444@mnnu.edu.cn

Received 31 March 2022; Revised 5 May 2022; Accepted 30 May 2022; Published 20 June 2022

Academic Editor: Mian Ahmad Jan

Copyright © 2022 Lanlan Cai and Zhiqiang Zhuang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The study of English and American literature is a significant part of curriculum for foreign language majors. In classrooms, multimodal theory teaching is often utilized, in which the single-modal interactive manner of textual or teacher’s vocal transmission is complemented and augmented by other multimodal teaching techniques, i.e., visual and auditory techniques. The information resources of English and American literature have dramatically expanded as a result of today’s information explosion and mature network technology. The lack of resources is not the bottleneck that prevents English and American literature from becoming studied. On the other hand, there are too many resources that depend on data mining technologies. This paper develops an integration analysis platform based on data mining for integrating the content of English and American literature resources and uses the optimized Apriori algorithm to process multichannel data. Using web technologies provides a single data management system that optimizes and queries the target data. We measure the quality of the target data of English and American literature resources and at the same time obtain the grey correlation degree of the data by calculation and assign a value function to find. The proposed method can optimize the target data by calculating the grey correlation degree of the acquired data and assigning a value function to it. The contribution value of the target data and the similarity discrimination threshold are obtained, which are used for calculating the total value. The simulation outcomes demonstrate that the proposed method can effectively simplify the data structure, enhance the data accuracy and quality, improve the query efficiency, and has high applicability and high-quality robustness.

1. Introduction

With the increasing spread and development of digital information and communication technologies, as well as the use of digital media in English language teaching (ELT), a more expansive understanding of what constitutes literature has become essential. Reading a book and watching a video used to be associated with different learning objectives, but that perception has changed in recent years [1]. When we read a book, we may suddenly be confronted with images, symbols, or other semiotic languages; in other words, we must decode semiotic signs in terms of making sense of the information. The scholars believe that literature has cognitive, entertainment, and compensatory functions. English and American literary resources are significant and useful for English learners, and their importance has been repeatedly emphasized by experts. In [2], the authors pointed out that, by reading literature, students learning a foreign language can learn how the speakers of that language approach the language and how they approach the language written in different linguistic contexts forms and conventions of symbols [3]. The literature explores the eternal and fundamental problems of human beings, which can transcend time, space, and culture and enable people to understand and restore the life scenes, cultural ideas, customs, and habits of a certain time in a country in an all-round way. Literary reading can enrich people’s life experiences and survival wisdom, sublimate people’s spiritual realm, improve moral quality, and purify the soul. Readers not only get pastime experience and physical and mental rest but also get an alternative satisfaction spiritually when they wander outside the world [4].
However, the shift of English education in universities and colleges towards practicality and light on humanities, the influence of pragmatism and utilitarianism, etc., have made the current study of literature cold, and the "uselessness theory" of literature study is very much in the air. There are several problems in the teaching of English and American literature courses [5]. British and American literature resource is the main course of English language and literature major. It is highly specialized and involves factors such as political economy, historical geography, religious culture, and other factors in European and American countries. It is difficult to teach and requires high knowledge of students. Because English and American literature itself has a long history, a variety of genres, abstract and hazy, and multiple perspectives, it is never easy [6]. The teaching model of a single text form leads to low student initiative. Due to the innate complexity and profundity of English and American literature classes, students generally feel difficulty in learning and serious intimidation. Teaching resources are of mixed quality. The bottleneck that restricts the teaching of English and American literature is not lack of resources; on the contrary, it is too many resources [7]. Teachers and students are faced with a wide range of choices in teaching and reading, with a variety of literature textbooks, selected readings, tutorials everywhere, a more abundant network of literary resources, and free downloads. Readers waste time on nonclassical books and resources while ignoring the use of quality materials and drawings. Moreover, these vast resources are inevitably mixed with good and bad ones, and the truly excellent materials are rare and drowned in many shoddy textbooks and teaching aids. The negative impact of these teaching resources of varying quality on students should never be underestimated [8, 9].

To solve the core problems of multisource integration and efficient storage of big data, we realized the standardized integration of multisource data by generating standardized metadata and constructing corresponding data dictionaries according to the composition and characteristics of data; based on data integration, he conducted a study on big data storage optimization based on the Hadoop platform, proposed a hash bucket storage algorithm considering the correlation of electricity distribution data, and realized the centralized storage of correlated data based on MapReduce to improve the efficiency of later data query and processing. The research proposes a hash bucket storage algorithm considering the correlation of electricity distribution data to realize the centralized storage of correlated data, improve the efficiency of data query and processing in the later stage, and realize the parallel correlated query of multisource big data based on MapReduce [10]. The authors proposed a prediction model and an algorithm integrating a moving average autoregressive model and a genetic particle swarm algorithm optimized wavelet neural network from both model construction and algorithm aspects. In terms of model construction, the first three moments of historical data with the strong correlation between the mean autoregressive model prediction and the grey correlation coefficient greater than 0.6 are used as the wavelet neural network input, and the model structure is simplified while taking into account the smooth and nonsmooth historical data, and the initial values of the parameters of the wavelet neural network are optimally selected by the genetic particle swarm algorithm, which can make the results less likely to fall into local optimum conditions [11, 12]. The authors of this paper discuss the most recent efforts to develop and test a novel multimodal process data capturing method. A mining technique to fusion provides a solid basis for automated production and the advancement of simulation models that accurately simulate building operations. To contribute to the identification of operational information needed to develop accurate simulation models, different types of operational data were collected and combined. The suggested algorithms are evaluated using laboratory-scale experiments, and the analytical results are also presented. The most important contribution of this research to the collection of information is that it provides the platform for further investigation. To build or improve, it is important to consistently identify the computer-interpretable right to information from varied data collected. Models from complicated, unstructured, and infrastructural projects, such as major construction and developing operations, are used to make accurate simulation models [13].

The above perversely literature methods need to extract data dimensional characteristics, which are computationally large and complex, and inefficient to effectively deal with massive data. Therefore, this paper proposes an optimization method for natural resources big data integration based on data mining technology which is based on the Apriori algorithm to process integrated multichannel data and optimize query processing to reduce processing steps and calculate the total value of data to complete optimization according to the grey correlation between data. Therefore, this paper proposes an optimization method based on the data mining technology. We propose a method of integrating and optimizing big data of English and American literary resources based on data mining technology. A data mining analysis system is built based on the heterogeneity of the big data of English and American literary resources for the content to be integrated. To achieve the target data integration and optimization, the Apriori algorithm is utilized to process data from different channels, and the total value is computed by evaluating the grey correlation degree.

Following are the main contributions of the proposed work:

(i) Firstly, to develop an integrated analysis platform based on data mining for the integration of the content of English and American literature resources and analyze multichannel data using the optimized Apriori algorithm.

(ii) To propose an optimization method for natural resources big data integration based on data mining technology, which is based on the Apriori algorithm to process integrated multichannel data and optimize query processing to reduce processing.

(iii) To improve the overall performance of English and American literature, data mining approaches have been frequently used to determine foreign languages' actual data of English literature using large amounts of multimodal data.
(iv) To develop the multimodal English and American literature resources, various evaluation approaches are employed to confirm that the data are analyzed using a mathematical statistics methodology and a data mining algorithm.

(v) An SQL and a T-SQL database built on the Visual Basic platform was used to collect Apriori algorithms and produce data mining rules for this study. Data have been integrated with the Apriori model and algorithm development technique.

The rest of this paper is arranged in a logical order: Section 2 expresses the related work, Section 3 highlights the pedagogical theory of multimodal teaching mode, Section 4 reflects the methodology of big data analysis and integration, and Section 5 depicts experimental results and analysis. Finally, Section 6 illustrates the research to a conclusion.

2. Related Work

It focuses on students using visuals and network pictures to mobilize and regulate the functioning of listening, perception, and other sensory functions. When movies, videos, sounds, and other elements are incorporated into the classroom, students’ perceptions of learning and writing improve, and the importance of art material increases. We can also increase the classroom teaching impact by reducing the boredom and obsession that a single text mode can contribute to the knowledge of literary works [14]. This study introduces the multimodality concept for rethinking ESL learning and teaching. It aims to look at its impacts and students’ perceptions of the usage of multimodal techniques, which are at the heart of the multiliteracies philosophy. The efficacy of multimodal methods to education insignificance, among 15 students in a private school in Penang, Malaysia, is investigated in this qualitative case study utilizing focus group interviews. The findings illustrate the importance to refocus English learning and teaching with a focus on multimodal pedagogical techniques to improve student learning outcomes [15]. This research examines wearables in depth and considers college English teaching as a potential future career. Teachers and students were polled using a survey questionnaire. The data from the survey are analyzed in this publication. According to the conclusions of the study, wearable technology can increase academic attention in college English. The goal of this work is to do multidisciplinary research on multimodal semiotics and humanistic teaching methods. The themes broaden the scope of the investigation and, to some degree, enhance the vitality of multimodal discourse analysis [16].

3. Pedagogical Theory of the Multimodal Teaching Mode

In a completed communication product or communication activity, the multimodal teaching mode refers to a blend of several semiotic modalities. But it may also apply to the multiple ways in which various semiotic materials are brought together to make meaning in a text. This theory is mainly based on social semiotics and Hanley’s functional grammar and suggests that symbols such as images, colors, music, and actions are traditionally considered secondary languages. In modern social communication, they are no longer considered an afterthought but rather a part of a larger semiotic resource that includes textual symbols and is used to generate meaning [17].

To engage students’ multiple senses in language learning, multimodal teaching advocates the use of multiple channels and instructional instruments such as the Internet, visuals, and role-playing. Students are encouraged to use as many of their senses as possible to engage in the connection between learning and teaching. To motivate and interest all students, we utilize the Internet, multimedia, and other resources, as well as visual, auditory, and tactile means, to the extent that teaching conditions allow. Multimodal instruction employs a variety of modalities or combinations of modalities, including visuals, colors, sounds, and words. Modalities or combinations of modalities include textbooks, literature, the Internet, movies and television, and music. The most significant criterion for modality selection in multimodal teaching is whether it can increase teaching efficiency. As a result, any teaching modality that can offer teaching situations and facilities as well as supply supplementary circumstances for teaching might be employed. The advancement of network technologies has resulted in perfect situations and conditions for learning and teaching. Multiple media engagement is an excellent technique for students to acquire course content and improve their learning.

3.1. Characteristics of English and American Literature Resources and the Feasibility of Multimodal Teaching

3.1.1. Situational. English and American literature resources include selected fragments of British and American literary history, literary scholars, and their works. They all focus on building a wonderful picture for readers, whether it is literary history, literary scholars, or literary works. Through their imagination, readers may see the Garden of Eden that God created for Adam and Eve. Enter Prince Hamlet’s palace, then proceed to the island where Robinson Crusoe lived alone, where you can observe the old man Santiago capture the massive marlin that was eaten by a shark. This sense of image is the expression of the situational nature of the literature course.

3.1.2. Plot. The writers’ works and characters in English and American literature all tell a touching story. Readers will sense as though they are walking down a long river of English and American literature as they read the works. They will have the impression that they are watching a variety of colorful characters complete various missions or face lots of new dangers, including the hero Belle Wulf, who was slain. Faust, who sold his soul to the devil; Hamlet, Adam, and Eve driven out of the Garden of Eden by God because they could not resist temptation; Crusoe, the adventure hero Gulliver; Tom Jones the outcast, Oliver the Mist West, and Tess, who was destroyed to death by society. Each of the characters and their stories in the work contain a touching plot.
3.1.3. The Feasibility of Multimodal Teaching Based on Data Mining Technology. Whether it is the story or the picture depicted in a literary work, when combined with the language of the characters in the picture, it gives life to the work and gives the reader a sense of immersive life. In this way, each literary text can integrate multiple modes of communication such as sound, text, image, and color, making the original single text mode multimodal. The computer network makes it possible to teach English and American literature using a variety of modalities such as voice, text, images, and even various colors and fonts and using students’ senses of hearing, sight, touch, and other senses to promote organic integration of language, pictures, sound, text, and other methods and symbolic resources in the literature curriculum [18].

3.2. Big Data Platform. The methods of data mining include fuzzy mathematical methods, visualization techniques, knowledge discovery methods, statistical taxonomy methods, and biomimetic methods. The application of these methods in distance education is still in the research stage.

3.2.1. Knowledge Discovery Methods and Applications. The decision tree approach, rough set method, and association rule analysis are the most common uses. The association rule analysis method assumes that \( I \) is a set of things and that each transaction \( t \) in a transaction database \( D \) is a nonempty subset of \( I \). Each transaction has its own TID (transaction identifier). The percentage of transactions in \( D \) that contain both \( X \) and \( Y \) is the support of association rules in \( D \). The percentage of transactions that include \( X \) and contain \( Y \) is the confidence level. The percentage of transactions that contain \( X \) and \( Y \) is the support association rules in \( D \). The application of the association rule method in distance education gave a complete data mining system design and implementation process and improved the Apriori algorithm to improve the efficiency of mining [19]. In the practical application, the mining results are used to guide the students to select courses and help them to better complete their assignments. Students can understand which courses to take first as basic courses when choosing a certain course, which is helpful for students to make scientific course selection and study. This study applied the Apriori algorithm to mine the students’ answer results in the examination system. Therefore, the association rule method focuses on the algorithm itself, and less consideration is given to whether these algorithms are suitable for the mining of distance education resources [20].

3.2.2. Statistical Taxonomy Methods and Applications. The application of statistical principles in data mining is relatively extensive and has formed an independent discipline. It mainly includes statistics, regression analysis, analysis of variance, and clustering methods. The clustering methods are divided into divisional clustering methods, grid clustering methods, density clustering methods, and combined grid density clustering methods. The authors proposed a clustering method based on association rules and hypergraph partitioning to cluster web pages and users of distance education websites [21]. In the application of distance education, some scholars have also studied that distance education learners need to be assessed by listening to lectures, assignments, and discussions to exams as a whole set of the learning process. Using the K-means clustering algorithm to cluster and analyze the recorded multiple assessment indexes and link the assessment indexes with the summative evaluation of learners overcomes the defect that the parameter indexes of the process assessment evaluation system are usually set based on experience [22]. The shortcomings of the process assessment system are usually set based on experience. Simultaneously, in distance education, personalized learning reduces redundant information of student attributes’ data. The rough set theory may also be used with a clustering technique, primarily to address the K-means clustering algorithm’s sensitivity to duplicated attributes in learner characteristics’ data that can deal with noisy data better and increase the clustering algorithm’s efficiency, making it suitable for distance education personalized teaching [23].

3.2.3. The Core Idea of These Methods. These methods are predicated on the concept of simulating particular behavioral activities of biological things using computer programs, particularly genetic algorithms and neural network methods. Some researchers used genetic algorithms to enhance the search engine and resource scheduling to increase the effect of finding and sharing network teaching resources so that students and teachers can get high-quality educational materials during the teaching and learning process. The fuzzy back propagation neural network method is utilized to build submodules such as FBPNN learning, knowledge base, and FBPNN inference machine for the university’s experimental formative evaluation process. The fuzzy neural network module is responsible for collecting, storing, and solving knowledge. The normalization of knowledge representations and expression conversion is achieved by input and output pattern conversion. It can better implement the university’s summative evaluation, reduce the interference of human uncertainty, and improve science and standards in the evaluation application. The BP neural algorithm is utilized in the new distance education teaching quality evaluation system to produce mathematical models with input due to several advantages of the neural network. Different evaluating indexes, output teaching effects, and scientifically and accurately evaluate the teaching quality and effect of current distance online education. To better solve the real-life problems and highlight their respective advantages, a novel research field of evolutionary neural networks was formed by combining genetic algorithms and neural networks, and many valuable conclusions and results were obtained [24].

4. Methodology of Big Data Analysis and Integration

4.1. Analysis of the Heterogeneity and Integration of Big Data of English and American Literary Resources. Due to its
special characteristics in data collection, representation, and management, natural resources big data has the characteristics of the multisource, multiscale, multidimensional, and spatial network in addition to the general characteristics of data.

4.1.1. Natural Data Collection. Natural data collection means and information content are multisource, with a wide range of spatial data sources and a variety of data collection means and methods, including field collection, remote sensing, digitization, and available statistics. The information content includes basic resource framework data, site base data, forestry resource information, meteorological information, and basic land resource information [25].

4.1.2. Data Characteristics and Storage. Resource space should have unit geometric characteristics and spatial relationships with different data characteristics and storage methods. Data preservation for different spatial data attributes, different storage formats and extraction methods, and big data formats are not uniform.

4.1.3. Spatio-Temporal Analysis Methods. Multispatial and multiscale: according to application requirements and data expression accuracy, spatial-temporal data exist in various size spatial units with significant spatio-temporal features. Due to the multisource and heterogeneous nature of spatial data, it is difficult to directly meet the cyberspace sharing and service requirements in the application process. As a result, a data standard and synthesis system must be established to synthesize and transform spatial data while maintaining a consistent spatial framework. We should create a unified data source for aggregation and sharing. The most challenging aspect of developing a master database is spatial data integration. Spatial data in different databases have different data formats, coordinate systems, data precision, and spatial time scales. Therefore, spatial data need to be extracted, transformed, and loaded to integrate them into an organic whole and achieve complete transformation or information exchange of multiple data sources [26].

4.2. Natural resource big data integrated analysis system. The design follows the data of reliability, efficiency, and practicality, and the processing process is divided into three stages according to the requirements of natural resource big data. The data collection layer, data mining, storage layer, and data analysis and display layer, and the structure of the system framework concept are shown in Figure 1.

Because the data collection layer recognizes that the generation channel differs from the data source, the platform can accept data supplied through various channels. When directly accessing the database and importing it into the data pool, SQL provides numerous interfaces for Office Automation Software (OA) access, while directly accessing the database and importing it into the data pool. The data from the other interfaces are transformed into excel data tables with explicit data structures using the visual basic application (VBA) language [27]. The data mining and storage layer, which fills in the data mining tables according to the attributes of the target information, constructs a new database. Since the integration of natural resources, big data studied in this paper does not need to face particularly complex unstructured data. Transactional structured queries based on the MSSQL database language (T-SQL) mining function is used to achieve mining quickly and perfectly.

When analyzing data and displaying the layer, a browser/server (B/S) architecture is used, which is suitable for remote access to different platforms and systems and enables a good graphical user interface (GUI) for customer-oriented data. In addition, the architecture is responsible for data analysis functions, such as creating a data model to analyze the comparison time of different data batch metrics over a fixed time [28].

4.2.1. The Data Integration and Analysis Platform. The platform extracts the source data into the SQL database based on the row and column relationships. So, this paper uses the T-SQL language to collect data in the SQL server database and recreate the corresponding database based on its results, as shown in Figure 2, the flow chart of data mining.

In addition, the goal of this layered design is to reconstruct the data extracted from the tables and create a well-structured database for further analysis and processing of the data.

4.3. Apriori Algorithm Based on Association Rule Optimization. According to the above description, the data association system is described in detail as follows: if we let \( I = (I_1, I_2, \ldots, I_m) \) denote the set of each data item, then we can know that the data \( D \) associated with the task is the set of database items, in which all event data sets \( T \) are nonempty subsets of the items so that we can obtain \( T \) [29].

The association system is in some way an implication of \( A \). If the calculation knows that \( A \) and \( B \) are, respectively, the set of data items. Then, we have \( A_1, B_1, \) and \( A \cap B = \emptyset \), and the result determines whether the support is the percentage of the data in \( D \) to the total event data, which is a sentence that can be written as

\[
\text{support} = \frac{\text{number of transactions containing } A \cup B}{\text{number of transactions containing } A},
\]

4.2. The confidence level is the ratio of the number of transactions containing \( A \cup B \) in \( D \) to the number of transactions containing \( A \) and is denoted as
condidence \((A \Rightarrow B) = P(A \cup B)\)

\[= P(A | B) = \frac{\text{support}(A \Rightarrow B)}{\text{support}(A)} \]  \hspace{1cm} (2)

Lifting degree:

\[\text{lift}(A, B) = \frac{P(A \cup B)}{P(A)P(B)} \]  \hspace{1cm} (3)

It can be derived from the above equation that the data sets \(A\) and \(B\) are both item sets with more frequently transmitted data, and \(D\) represents a large set that contains all item sets of the form \(A\), and \(BP(A|B)\) and \(P(A \cup B)\) are then described as the conditional likelihood of data mining and the quantitative item set and likelihood, respectively [30].

According to the above formula, the higher the value, the higher the degree of association between the data, and the main purpose of using data mining methods is to obtain frequent item sets with high support and confidence. The efficiency of integrating data can be greatly improved by using the mined data item sets. In the subject of association identification, the Apriori algorithm is a traditional method that can be implemented in two ways. Using the database to generate association rules that meet the minimum trust requirements: iterative computation for the common items in the event data set and establishing them as the minimum support threshold based on specific outcomes [31].

In this paper, T-SQL language is used to obtain the parameters of association rules and to determine the data mining rules. The specific optimization of Apriori modeling and algorithm implementation process is as follows.

4.3.1. Calculate the Global Frequent Item Set. Since the integrated data are not a very complex structure, there is no need to search for the global common elements in a haystack, as is usually done for data extraction from the Internet.

4.3.2. Obtaining Strong Association Rule Degrees. Obtaining semantic information, location, and data structure characteristics of the target data to be mined and constructing relational rule tables using reference items.

4.3.3. Creation of Association Rules. For the creation of association rules, multiple scans of the data pool are not required.

According to the analysis of the above algorithm, the integrated data of natural resources can be easily obtained. The data form can be analyzed and processed, and the resource data and the data of various indicators that need to be viewed are directly displayed.
4.4. Natural Resources Big Data Integration Optimization Implementation. In many complex data, first, we should define the dependency matrix between the target data and then distinguish the attributes of each data set to get a set of data attributes. Then, we need to perform Fourier transform on these data attributes; Fourier transforms the data attribute set and analyzes the data attribute ratio, vector similarity calculation, and demand optimization.

4.4.1. Calculation of Correlation between Different Data. The D-S evidence theory is used to measure the quality of single and multisource data and the reliability of the target data category and to calculate the correlation coefficient between the data.

Suppose \( I = (I_1, I_2, \ldots, I_m) \) represents the m-dimensional target data, \( p_{str} \) represents the completeness of the target data formed by one data root alone, and \( F_{set} \) represents the similarity of multiple target data, such that the data quality can be calculated using the following equation:

\[
M'_{str} = \frac{F_{set} + S_{i1} + S_{i2} + S_{i3}}{P_{dr} \times m + \varepsilon_{str}} \times v_{str}.
\]  

Here, \( p_{str} \) is the number of data tuples generated by the data sources, and \( v_{str} \) is expressed as the likelihood function [32].

If \( \partial_{d1}^i \) is used to represent the basic probability magnitude function of the evidence provided by \( i \) individual data sources, \( c_{dyi} \) is the likelihood magnitude in the categorical evidence. The target data have to be queried, and \( R_{dtr} \) represents the difference between the precise categorical likelihood take of the data category and the other categorical confidence take. Then, the confidence of the different target data belonging to their categories can be obtained according to the following equation:

\[
M_{su} = \frac{R_{dtr} + \partial_{dghjk}}{n + \partial_{dghjk}} \times S_{op}.
\]

Here, \( S_{op} \) is expressed as the vector of weights with the highest likelihood values among different data categories.

Based on the above formula, assuming that \( \varepsilon_{fryu} \) is described as the reference column of the data to be integrated, and \( \mu_{str} \) and \( \phi_{str} \) are expressed as the number of indicators and data series, respectively. Then, the correlation degree between the data columns can be obtained by the following equation:

\[
B'_{depp} = \frac{\phi_{str} + \mu_{str}}{\varepsilon_{fryu}} \times M'_{su}.
\]

In summary, it can be shown that the credibility of each data category is indicated in the case of diverse needs for the optimization of heterogeneous target data in IoT. The degree of correlation between data columns is measured by the grey correlation theory. The correlation coefficient between the individual data is obtained, which lays the foundation for realizing the optimization query of multisource heterogeneous target data in IoT.

4.4.2. Optimization of Heterogeneous Target Data Integration Based on the Total Value of Similarity. Based on the above-derived correlation coefficient \( v_{str} \) of each data as the actual basis, the dynamic weighting method is applied to select the data attributes with obvious differences. Then, the similarity is obtained by calculation, and the target data to be integrated are optimized according to the results, where the specific process is as follows:

Assuming that the key attributes of any of the target data are represented by \( \mu_{gyu} \) values, equation (7) is used to select the key attributes of the differentiated data according to the correlation coefficient \( v_{gyu} \) of each data.

\[
R_{cy} = \frac{N + \mu_{gyu}}{S_{gyu} + W_{gyu}} \times v_{gyu}.
\]

where \( W_{gyu} \) represents the discrete type of the key attributes of the data [28].

Because the large data of natural resources contain overlapping subsets within it, it can be considered as \( P_{gyu} \), while using \( X_{i0} \) described as the \( i \) attribute of \( k \), and then the entire data can be divided into different sets based on this attribute, i.e.,

\[
R_{r1} = \frac{i + X_{i0}^t}{k_{j1} + W_{j1}} \times P_{gyu}.
\]

According to the result of the above formula, if we let \( \partial_{gyu} \) denote \( f_{gyu} \) the attribute of a certain set \( r_{hjk} \) taking values,
then we can know that $x_{jg}'$ and $y_{jg}'$ denote $k_{xb}'$, the attribute of two random sets $\theta_{jg}$, as well as $k_{jg}'$, so that we can write to calculate the attribute similarity as follows:

$$W_{wqe}' = \frac{\partial_{jg}'}{r_{jk}d_{jg}} \sum_{z_{ji}} \left[ \frac{\partial_{jg}'}{d_{jg}} \times \frac{\partial_{jg}'}{2} ight] a_{s_{xb}}' \pm k_{xb}'$$ (9)

where $z_{ji}'$ represents the weighted average of the attribute similarity of each data set, while $a_{s_{xb}}'$ is described as the proportion of contributions from the attributes of different data sets.

If $\lambda_{de}'$ is used to express the weight value of the integrated set attribute $p_{lmo}$, and $\partial_{jg}'$ denotes the sum of all attribute weights, then it can be described as the actual availability of the attribute $p_{lmo}$ as follows:

$$e_{lmo} = \frac{\lambda_{de}'}{\partial_{jg}' \times h_{lmo}}$$ (10)

where $h_{lmo}$ represents the maximum value of the weights of each attribute.

Assuming that the effect of incomplete records of data attributes represented by $I_{jk}p$ on the similarity calculation, the target data integration is performed using equation (11) to optimize the query.

$$v_{lmo} = \frac{\lambda_{de}'}{e_{lmo}'} + R_{cy} + W_{wqe}$$ (11)

According to the above steps, the optimization of IoT multisource heterogeneous type target data is completed, and in this way, the optimal query effect is obtained.

5. Experimental Results and Analysis

A comparison with the literature is needed to evaluate the suggested method’s practical efficiency (method 1 in Figure) even further and (method 2 in Figure) is performed. The items of comparison are the efficiency as well as the accuracy of the input directory to find the corresponding data after integration and optimization. In the actual simulation experiments, the three methods are used for query optimization of target data on the Internet to determine the error rate of the query to be optimized, and the experimental comparison results are shown in Figure 3. According to the experimental comparison graph in Figure 3, the technique in this study can efficiently handle the relationship between heterogeneous target data from multiple sources and ensure the quality of target data by reducing the duplication and redundancy of data, while the literature method has a lower error correction rate compared with the method in this paper. The method conducts data integration optimization experiments using 140 data sets for 14 experiments, and the results obtained by the three methods in different experiments are shown in Figures 4 and 5.

As shown in Figure 4, the integration efficiency of the method in this paper is relatively high because the weighted average of the calculation and data attribute recognition can be effective. We need to distinguish the contribution rate of different sets of attributes and derive the actual availability of data to obtain a more integrated effect.

Figure 5 is compared with the literature method [4], and the dynamic weight assignment method is more scientific. In this paper, we can optimize the natural resource big data query, fully meet the integration and optimization needs of the target data, and fundamentally improve the efficiency of the integrated and optimized data query.

From Figure 4, it can be seen that the integration efficiency of this method is relatively high because the weighted average of the calculation and data attribute recognition can effectively distinguish it. The contribution rate of different sets of attributes is derived from the actual availability of data to obtain a more integrated effect.
6. Conclusion

The multimodal discourse analysis theory has been utilized in teaching as a new teaching model, addressing the limitations of conventional teaching, reforming restrictive educational theory, and adjusting to educational development. At the same time, today’s environment requires education to be adapted to each student’s specific requirements. Due to the problems of low query efficiency and ineffective integration of massive data by traditional methods, this paper proposes a data mining technology-based multimodal English and American literature resource integration research optimization method. According to the heterogeneity of the big data of English and American literary resources, a data mining analysis system is constructed for the content to be integrated. The optimized Apriori algorithm is applied to process the data of different channels. The total value is obtained by calculating the grey correlation degree to achieve the optimization of the target data integration. During experimental processing, it was determined that more accurate results are required for computing the correlation degree. The error threshold should not exceed a particular range; otherwise, it will have a significant influence on the outcomes, requiring further optimization of the algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.
Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Horizontal Project of Shanghai Foreign Language Education Press Co., Ltd. “The Elucidation of Feminist Theory in English and American Literature” (2020SH49).

References


