Electronic educational content formation with the application of cloud technologies

Firudin T. Aghayev¹, Gulara A. Mammadova², Esmira R. Alasgarova³

¹²³ Institute of Information Technology, B. Vahabzade str., 9A, AZ1141, Baku, Azerbaijan

¹agayevinfo@gmail.com; ²gyula.ikt@gmail.com; ³aleskerova-66@mail.ru; ⁴orcid.org/0000-0003-3685-6145, ⁵orcid.org/0000-0002-6332-7129, ⁶https://orcid.org/0009-0006-0719-7418

ARTICLE INFO

http://doi.org/10.25045/jpis.v14.i2.10

Article history:
Received 8 February 2023
Received in revised form
11 April 2023
Accepted 13 June 2023

Keywords:
E-learning
Cloud technology
Educational content formation
Requirements
Data mining techniques
Software

ABSTRACT

Many experts and scientists focus on cloud computing, mobile Internet technologies, and the rapid development of the Internet of Things in the field of e-learning which caused innovation and revolution. The development of the capabilities of cloud technologies has been applied to solve problems in various fields of human activity, including the educational process. Cloud technologies enable students to actively participate in the preparation of educational content, modify learning methods, increase access to information resources (external memory, applications, educational materials, etc.). Clouds unify physical or virtualized resources in data centers and require less software and hardware. The goal of this research is to expand and use the possibilities of applying cloud technologies in the teaching process in higher education institutions. As a result of the conducted research, we determine didactic and information opportunities by using cloud services in the educational process. This article reviews scientific research conducted in the field of data mining and cloud technologies in the formation of educational content during the last decade. Moreover, the methods of solving educational content management problems are proposed using data mining methods.

1. Introduction

Technological innovations brought by Industry 4.0 penetrate all areas of human activity transforming the activity of society as a whole. Artificial intelligence, robotics, Big Data and the Internet of Things have a significant impact on education as well as on other fields, as the labor market, economy, medicine, etc. Education 3.0 has been left behind being replaced by Education 4.0. With the arrival of Education 4.0 to educational institutions, the concept of education is completely changing, new trends are emerging. Due to new interactive learning tools, education is no longer dependent on time and space, students have the opportunity to study anywhere and anytime. All educational opportunities of information technologies are used in the most extensive and comprehensive way in the electronic education system.

Recently, the increase in the flow and volume of data creates problems related to their storage and processing. The amount of information that needs to be mastered independently in e-learning is constantly increasing. Moreover, new methods for their processing and systematization are developing. Cloud computing is a new approach in data collection, processing and transmission. As one of the promising fields in ICT development, cloud technology can be used to solve many problems in the educational process. The application of cloud technologies in education not only reduces the cost of the educational process, but also increases its efficiency (Kononyuk, 2018).

Cloud technology (cloud computing) refers to distributed data processing and provision of
necessary services through the Internet. Higher education institutions practice three main cloud models, i.e., Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS), and Platform-as-a-Service (PaaS) (Arunachalam et al., 2014).

IaaS offers on-demand hardware provisioning (virtual machines, servers, data storage devices, networks, firewalls, IPs, etc.). PaaS provides computing platform for building hardware or software applications without obtaining a license. Typical services provided include operating systems, operating environment, database, web server, working tools. SaaS ensures users with access to software and databases without installing them on their devices. Cloud service providers manage the infrastructure and platforms on which applications run. This is often determined on payment and usage. The structural scheme of a cloud-based higher education institution is presented below (Figure 1).

![Diagram of cloud-based higher education](image)

Fig. 1. Architecture of cloud-based electronic university (Seydiametova, 2012).

Educational management systems such as Microsoft Live@edu and Google Apps Education Edition, which are the most popular cloud services in educational practice, Windows Azure, Moodle, WebCT, E-Learning Server, Prometheus, etc., serve to promote social collaboration, as well as provide an opportunity to improve communication and partnership. They provide access to Google Art Project information resources, Google Earth and National Geographic multimedia services (Ananchenko, 2015).

Like all technology, cloud technology also has both advantages and disadvantages. Advantages of cloud computing are as follows:

- User pays for the service only when he/she needs it, and most importantly, he/she pays only for what he/she uses;
- Cloud technologies make it possible to save on the purchase, support and improvement of software and equipment;
- Scalability, fault tolerance and security – automatic allocation and provision of necessary resources depending on the application needs.
- Maintenance, software update is performed by provider;
- Remote access to data in the cloud - accessing the Internet from anywhere.

Disadvantages of cloud computing may include:

- User does not own the cloud resources and does not have access to the internal cloud infrastructure. Security of user data is highly dependent on the provider company;
- A shortcoming that is relevant for Azerbaijani users: in order to obtain quality services, the user must have reliable and fast Internet access;
- Some data cannot be entrusted to the provider not only for storage, but also for processing;
- Not every program, for example, allows saving the intermediate stages of data processing, as well as the final result of the work on a flash card, and online results are not always convenient;
- Possibility of a risk of data loss due to failure of the online service provider to backup and server crash;
- By entrusting the data to an online service, the control over them is lost. The user cannot change some of his data, they are stored in conditions independent of him.

Nevertheless, cloud technologies allow students to overcome geographical and socio-territorial barriers, to get the necessary textbooks and other educational materials in a more optimal and reliable way. Distributed cloud computing is one of the effective tools for solving data processing problems in education (Berman et al., 2003). Cloud computing platform provides distributed data storage and parallel computing capabilities. Data mining in cloud computing provides the process of extracting data from unstructured or semi-structured web data sources.

This article offers solutions to the problems of educational content management using data mining methods in cloud environment.
2. Related Work

The application of mining methods in education is realized in 3 trends: descriptive analytics, predictive analytics and prescriptive analytics (Ogurtsova et al., 2021).

Descriptive analytics focuses on the description and analysis of historical data collected about students, faculty, research, courses, and other administrative processes. The aim of such analytics is to enable higher education institutions to analyze information about, for example, the access frequency, page views, program registration, transactions and interactions in the process of education and scientific activity, and to estimate the ratio of successful and unsuccessful students in a certain course.

Some works show that there are a number of cloud services for educational institutions that help students gather the necessary knowledge and conduct interdisciplinary and complex research. For example, Ramesh A. et al. (Ramesh et al., 2021) describe a neural network called DALL-E that generates images from text headings for a wide variety of concepts expressed in a natural language. The network connects a large number of real and imaginary objects and enables “thinking” what needs to be done to put them into practice.

At present, the rapidly developing cloud computing technology is gradually integrated into IoT (Internet of Things) data mining technology and forms a new model. Based on this, Ding, Bangxu et al., in their article (Ding et al., 2019), explore the use of IoT data mining based on cloud computing technology.

Predictive analytics focuses on assessing the probability of future events and determining their occurrence risk by examining trends and identifying associations. Predictive analytics can uncover hidden relationships in data that are not visible in descriptive analytics, such as demographics and course completion rates. This analysis can be used to pre-identify students who are most likely to pass the course based on their performance in the first weeks of training (Morsy et al., 2017). Analytics help instructors track each student’s course completion rate and examine how course tools and content are related to student success.

Guided analytics helps universities assess the current situation and make informed choices about a preferred course by calculating expected future outcomes and alternative solutions (Polyzou et al., 2016).

Many learning algorithms applied in data mining are either supervised learning or unsupervised learning. Supervised models are sometimes called predictive models. These models predict the target value. A supervised model in cloud computing is a classification model.

The classification of the collection refers to grouping or categorizing the elements that make up the collection. In the data mining context, classification is performed using a model based on historical data (Kabakchieva, 2013). The predictive classification aims to accurately predict the target class for each of the new data, i.e., the data not present in the historical data (Bunkar et al., 2012).

Classification in cloud computing often uses the following methods or algorithms: k-means, decision tree, support vector machine, etc.

Moretti C. et. al propose a strong abstraction called Classify to group sequential programs into parallel data graphs (Moretti et al, 2008) and enable them to be deployed seamlessly in clusters, clouds or multiprocessor machines. In their study, they evaluate three different and well-known learning algorithms of variable complexity on databases of varying sizes up to 54 GB.

Another study (Nanjang, 2010) proposes an ensemble solution for a large number of models, each of which is used to review different aspects of the data. Among these different models, k-nearest neighbors (KNN) and restricted Boltzmann machine (RBM) are found to be successful models.

Some studies present a distributed algorithm to learn the Bayesian network parameters from distributed heterogeneous datasets (Lagman et al., 2019). This method spreads the database over multiple sites, and each site includes observations for a different subset of traits.

Unsupervised learning models. These models do not predict the target value, but focus on the internal structure, relationships and interactions of the data. Unsupervised learning models are sometimes called descriptive models. In cloud computing, the unsupervised model uses two data collection methods: clustering and association rule.

Mining of associative rules is the most common EDM technique. The relationship found in the analysis of association rules refers to “if” \( \rightarrow \) “then” rules. The main goal of these rules is to determine whether one event leads to another or not. This method establishes a relationship between two or more variables in a data set. In education, for example, it is possible to establish a link between parental educational attainment and
student dropout, or to clarify which pedagogical strategies lead to more effective/reliable learning (Oughdir, 2019).

Association rule in cloud computing commonly uses the following methods or algorithms: apriori algorithm, improved apriori algorithm, parallel association rule extraction algorithms, association rule extraction algorithm, sector/sphere and MapReduce.

The strategy of association rules in cloud computing environment is analyzed in (Polyzou and Kapuris, 2016). This article introduces cloud computing data algorithm, hadoop, mapreduce programming model, apriori algorithm and parallel association rules. Later, it presents a strategy based on parallel fusion rules adaptable to cloud computing environment. Using associative methods, the data processing technique applies advanced apriori algorithms for parallel analysis of association rules.

The clustering method is beneficial for exploring the data and discovering clusters embedded in the data. A cluster is, in a sense, a collection of similar data objects. A good clustering method produces high-quality clusters to ensure low similarity between clusters and high similarity within a cluster. The following methods or algorithms are commonly used for clustering in cloud computing: data indexing algorithm, k-means method, hierarchical virtual k-means algorithm (Pandagale, Surve, 2016).

Hadoop distributed file system. Several studies propose methods for using data indexing algorithm for multidimensional large-scale data sets. Gopalakrishnan N. and Madhuri K. use a hierarchical k-means method to solve the problem of data generation and refined data collection designed to collect from multiple data centers and multiple processors and to integrate relevant data (Oyelade et al, 2010). This technique is used for storage operations, SAAS, PAAS, etc. creates a clean output of homogeneous data distributed across multiple computers and data centers in the cloud.

3. Research materials and methods

Our study makes it possible to identify the main trends in the use of cloud technologies in e-learning in the era of Industry 4.0. In recent years, technological development has had a great impact on educational institutions, and the use of cloud services enables students to access more information, and consequently, more knowledge.

The development of Industry 4.0 lies on smart technologies, artificial intelligence, big data, robotics and cloud technology services. In this new environment, educational institutions face the challenge of training successful graduates and creating new learning methods.

During our research, we perform the analysis with the reference to the scientific publications from the databases of Springer Link, IEEE Xplore, ACM, Science Direct, Google Scholar, as well as eLIBRARY.ru. In the above databases, the following keywords are used as search parameters: “the use of cloud technologies in e-learning”, “the formation of educational content in the cloud environment”, “the use of Data Mining in the formation of educational content”. Scientific publications of the years 2010–2022 (including early 2023) are analyzed.

Initially, the search for articles is carried out using the Google Scholar search engine, which allows finding articles in a large number of scientific journals. Each database is further searched using appropriate tools.

The following common scientific methods are used in the course of the research: analytical review of the problem, synthesis methods, induction, generalization and comparative analysis.

The analytical review of the problem focuses on the personalization of the content of educational programs in higher education institutions and the importance of using cloud service infrastructure. The user needs to turn to cloud technology to get rapidly increasing and complex structured data for training.

The content of education and the main types of educational activities (basic and additional training, retraining, counseling, professional development) are formed taking into account the user categories (applicants, students, graduate students, specialists, employees). The infrastructure of the educational content and its management technology are determined by the methodology of the teaching process, the capabilities of the virtual information and communication environment of the educational institution. Therefore, the requirements for the formation of a unified educational space include the approaches to teaching and communication language, ensuring the continuity of knowledge management practice, the development of the modern corporate culture of the university, and the creation of educational content.

A set of elements of the university’s e-learning resources can be represented as:
M=(VD, LS, PS, BK, WB, NX) (1)

- VD - video lecture,
- LS - materials for independent work (for example, a lecture in pdf or html format),
- PS - lecture notes and slides,
- BK - teaching instructions,
- WB - links to additional resources (electronic libraries, maps, photos, any other social media resources and Internet resources),
- NX - supplementary text and multimedia materials.

A set that determines the activity of the student can be represented by the elements as follows:

F=(Q1, Q2, Q3, SM, Gl, GM) (2)

Where:
- Q1 - test or control questions,
- Q2 - tasks (report, essay, term paper, exercises, assignments),
- SM - electronic seminar on mutual evaluation of performed work,
- Gl - glossary,
- GM - educational games including simulations.

A set of e-course feedback elements can be presented as follows:

E=(FR, BL, SN, NS) (3)

Here:
- FR - a forum (discussions, consultations, online learning communities),
- BL - a blog,
- SN - social networks,
- NS - other forms of feedback.

The main didactic requirements for the educational content should be represented in 3 features:
- relevance of the active learning model;
- support for current independent work of students in studied subjects;
- ability to customize content infrastructure and understand individual topics.

Each training materials consists of training modules implemented in cycles, namely: educational activities, learning (lectures) and execution of theoretical materials, practical tasks. The number of training modules should be optimal according to the number of topics of the studied subject, as well as the set of minimum competencies to be formed by the end of the study of this subject.

During the research, we find out that cloud technologies stimulate practical interaction of university teachers in the process of solving actual issues related to the formation, filling and regulation of the main blocks of educational content and each subject (thematically organized content, types of stepwise current tasks for self-control, methodical recommendations and a set of materials). The use of cloud technologies in the interactive cooperation of representatives of various subject areas facilitates access to information resources and multimedia services, and enables interdisciplinary communication and research stimulation.

When using cloud technologies, the role of cooperation between university professors, computer scientists and scientific librarians increases in solving the problem of producing subject-oriented electronic resources, i.e., during the search, processing and inclusion of special resources in public electronic databases, their evaluation and optimal development.

Subject-oriented resources of educational content, that is lectures, practical studies, tests, tasks, reference lists, interpreted web links, must meet the requirements of openness, completeness and system mobility. These necessary features should be ensured by using cloud technologies as a mechanism for information and methodical provision of the virtual learning environment.

Cloud technologies in education require more use of cloud services not only on online platforms, but also in educational institutions for joint storage of teaching materials, scientific and creative projects. Moreover, cloud technologies control the knowledge of students: online tests and control exams are now performed even in the most remote boundaries of the country. Students and teachers are increasingly using cloud solutions for activities outside the classroom to organize classes. Today, on local cloud educational platforms, students can quickly find almost any information they need (Butko, 2015).

Organizational principles of educational content formation should be determined by communicative ideology (Shaunda, 2010). Cloud technologies stimulate the practical cooperation of university professors and teachers in the process of forming, filling and editing the main blocks of educational content for each subject.

In this regard, the educational content should be created as a distributed knowledge base that ensures a set of e-learning tools and educational information reserves, their coordinated and efficient use by all participants of the educational process.
4. Application of Data Mining technology in electronic education system

Educational content should be created as a distributed knowledge base that combines electronic educational tools and educational information resources, and also ensures their consistent and efficient use by all participants of the educational process. The basis of educational content lies in cloud storage, an online storage model on the Internet. The use of cloud technologies enables the data sharing between specialists on different disciplines, intensification of interdisciplinary relations and research. This allows solving the problems of accessing public electronic databases and search processing, developing methods for their effective use in the educational process, and creating electronic educational resources.

The electronic education systems of educational institutions have been collecting information about various aspects of the educational process for a long time, mainly about the students, their progress and attendance, teachers and their scientific, educational and administrative activities, educational content (text, audio, video), etc. This data must be efficiently stored, processed and analyzed.

However, the use of data mining methods associated with this huge database is limited because the amount of computing resources required to support an intelligent system with these functions in educational institutions is insufficient. In this structure, the use of cloud computing environment is mandatory. Due to the opportunities offered by cloud computing in terms of high-performance and easily scalable computing and storage resources, the use of EDM (Educational Data Mining) is becoming relevant (Romero and Ventura, 2013). Standard software and computer systems can lose many useful features, as they cannot process all this information in a reasonable period of time.

Data mining is the process of discovering useful information in a large database. Cloud platforms provide the user with various utilities, services and tools for processing and performing these analyses. Educational data mining is associated with the measurement, collection, analysis, and presentation of educational data to optimize both the teaching environment and the teaching itself. Broadly speaking, EDM software and methods are used to improve the organization of the educational process, to increase its effectiveness, to examine the student knowledge and educational institution data (Long, 2011).

Cloud computing provides a centralized platform for software and data storage to provide efficient, reliable and secure information services to users of educational institutions. (Dhope and Deshpande, 2016) Cloud computing provides mining software and hardware for data acquisition over the Internet as a service.

The main advantages of using data mining in obtaining data provided by the cloud are as follows:

- Users only pay for the data mining tools they need, which reduces their costs because they don’t have to pay for complex data mining kits they don’t use to their full potential;
- Users don’t need to maintain hardware infrastructure because it can search for data through a browser, which means that he/she only has to pay for cloud computing when he/she uses it.

Figure 2 illustrates the data mining tools provided to users of educational institutions by cloud services ((Educational Data Mining, 2023).
5. Software used for processing educational content in the cloud environment

This section presents software samples for the generation of educational information with successful use of cloud technologies in educational institutions.

Currently, cloud services provide students with a variety of services to build their own machine learning programs. In the educational process, they can use the following cloud technologies:

- **Makeblock** includes five artificial intelligence tools: smart services, machine learning, text-to-speech translation and vice versa (Vostinar, 2020).
- **PictoBlox** offers computer vision, face recognition, optical pattern recognition, language recognition, machine learning, ethics in artificial intelligence, the Internet of Things (Agilo Research, 2020).
- **Teachable Machine** is a Google machine service (Teachable machine, 2021). It can recognize images, voice commands and human actions.

All these services are free and available during classes. The only requirement is that the device has enough power to stream video. These resources create a fully functional program without deep knowledge of the mathematical foundations of artificial intelligence and programming. At this stage, students gain insight into the network learning process, understand the essence of datasets, and the size of a data sample. Consequently, students acquire the skills to build datasets for network training.

The next stage of training is the transition to programming. It is aimed at mastering the basics of algorithmizing and programming. Kaggle is an analytical and predictive modeling competition platform that can be used to support this phase (Kaggle, 2022), (Chow, 2019).

Another cloud service can be successfully used in educational institutions. This is Amazon Elastic Computer Cloud (Amazon EC2) service. It is a Hadoop-integrated, web-scale, web service with computing power in the cloud: runs on Map Reduce Linux, Open Solaris and Windows servers. Amazon EC2 is designed to ease web-scale computing for developers, and the web service’s simple interface enables to purchase and configure resources with minimal effort. It has full control over computing resources and enables working in a reliable computing environment (Saini and Behl, 2020).

Windows Azure is an open and flexible cloud platform that enables to rapidly create, deploy and manage applications in a global network of data centers managed by Microsoft. F# language is used to generate data in cloud computing. This environment allows to create applications using any language, tool or platform and integrate public cloud applications with available IT environment.

OpenStack enables collaboration between cloud developers and technologists producing a private and public open-source cloud computing platform. This technology aims to provide solutions for all cloud types that are easily deployed, scalable, and feature-rich. OpenStack consists of interconnected projects providing various components for a cloud infrastructure solution.

6. Conclusion

E-learning system users face the challenge of optimizing large-scale resource management due to the huge growth of services, educational content and media resources.

This article emphasized the relevance of using cloud computing in the field of education. The advantages of integrating an e-learning system in the cloud lie in high flexibility and scalability of resources, including storage, computing requirements and network access. We also highlighted both the software and hardware advantages of using cloud resources in education. Cloud services, on the one hand, provide various educational programs at a lower license price, and on the other hand, provide software updates.

The importance of using a cloud computing platform when using Data Mining methods in e-education was also underlined. The computational advantages of this new cloud computing paradigm lie in the extraction of useful knowledge through intelligent data algorithms and the efficient provision of the obtained data for educational institutions.

The use of Data Mining methods in the cloud environment ensures maximum adaptation to the modern requirements of students in electronic education. Taking into account the individualization and monitoring of the activity of each student, it not only improves the teaching materials, but also individualizes the process of teaching and cognitive activity.
This article analyzed the possibilities and perspectives of using cloud technologies to improve the quality of services and increase efficiency in the management of electronic educational content.

References


Educational Data Mining, 2023 https://www.javatpoint.com/educational-data-mining


