

Creating a distributed, privacy-aware e-learning system based on item response theory

Bojan Kostadinov
Software Engineering
Cloud Solutions LLC
 Skopje, Macedonia
 bojankostadinov@gmail.com

Irena Stojmenovska
School of Computer Science and
Information Technology
University American College Skopje
 Skopje, Macedonia
 irena.stojmenovska@uacs.edu.mk

Abstract—The quality of education is an important objective for governments and organizations around the world, as education has the potential to open new opportunities for people, improve their lives, and make them a productive and useful member of society. One of the ways to improve education is through the use of technology, and by making sure children have access to knowledge and information. In this paper, we will describe a distributed e-learning system based on item response theory, which can intelligently offer students appropriate questions and tasks to solve, and adapt the difficulty of the presented test items based on the students' previous answers and activity. Similarly, after each test item, the system presents the user with a solution to the problem, and the theory behind it. Our data and survey scores show positive results, and outline that students find the system engaging and want to continue to use it in the future.

Keywords—distributed systems, e-learning, item response theory.

I. INTRODUCTION

Governments and organizations throughout the world have recognized education as an important tool to reduce poverty, teach values, create a more productive and confident workforce, and to promote equality by empowering people from various underprivileged groups. Through education, people learn and acquire the necessary facts, principles, standards, norms and skills in order to be a successful, confident, useful and productive member of society. An educated workforce is crucial for economic growth.

The use of technology and e-learning is one of the main methods that allows us to bridge the gap between people that come from different backgrounds and socioeconomic status – i.e., to allow people to have similar access to learning opportunities. The ability of e-learning systems and technology tools to scale, to assist groups and individuals in learning at their own pace, to cover material according to an individual schedule, and to digest up-to-date content, are just some of the advantages of electronic learning. Therefore, it is very important to analyze their capabilities and usage in both formal and informal education. One of the current research areas in this field is related to the use of intelligent and adaptive e-learning systems, which have the ability to automatically adjust to students' capabilities [1].

In this paper, we will describe a new distributed, privacy-aware e-learning system based on item response theory, including data related to its usage for improving students' scores in mathematics.

First, in Section II, we will look at present research related to distributed systems, privacy, item response theory and existing e-learning systems. Next, in Section III, we will present the architecture of the newly created e-learning system, as well as its features. In Section IV, we will outline

data gathered from using the system in the course of one week, as well as results from a survey conducted among students who used the system during that period. In Section V, we will identify our ideas for future work - and finally, in Section VI, we will write a conclusion about our findings.

II. LITERATURE REVIEW

A. E-learning

The role that e-learning software plays in supporting and improving educational systems across the world continues to increase every year. This includes their usage in both primary schools, high schools, universities and informal settings like online courses and certifications. The global COVID-19 pandemic has expanded this growth even further, with some organizations and institutions even providing entire curriculums online. In some countries, like Macedonia, an entire school year was conducted online.

E-learning has several advantages compared to traditional learning, although in most cases they run in parallel – i.e., e-learning systems are commonly used to augment traditional learning. Blended learning is often considered to be a better and more effective alternative to both pure online classes and pure face-to-face classes [2].

E-learning systems offer students and teachers the ability to communicate and exchange messages asynchronously, reduces costs, and allows students to review and learn information at their own pace. Additionally, e-learning is scalable, (often) offers personalization features for the students, and analytics (data) for the teachers and administrators. This can lead to better learning outcomes, and better retention – as different students have different learning styles.

Learning management systems might be designed to handle various types of courses, or they might be targeting a specific area, market or expertise. For example, Moodle is often considered the most popular general-purpose Learning Management System [3], with some data showing that it has as much as 64% of the European LMS market share. It is an open-source platform, which is easy to setup and maintain. It offers a personalized dashboard for users, file management, calendar, notifications, multimedia, assessments, custom integrations (through plugins), and a lot more. From a technological point of view, it requires PHP and a working web server like Apache, as well as a relational database (MySQL, MariaDB or PostgreSQL).

An example of a LMS that is used in a specific area (computer programming) is MENDO [4], which allows students to read lectures which contain interactive elements, solve tasks with automatic grading, participate in competitions and forum discussions, etc.

Some systems, called intelligent e-learning or intelligent tutoring systems, are able to provide immediate and customized feedback to students, and can be utilized even when the physical instructor/teacher is absent [5]. Some examples of intelligent tutoring systems are those that can offer automatic and individual assistance, hints or adaptable task selection. In this paper, we are going to describe how our system automatically selects tasks for each student and enables them to learn and improve in the specific areas that are most important to them.

B. Distributed systems

Most e-learning systems are installed on a single web server, either with a relational database management system running on the same machine or in the same datacenter. For example, using Amazon Web Services, Moodle can be installed on a single EC2 instance, and then connect to a database running on RDS (in the same region). This kind of setup is very simple to configure, manage and has very little network overhead.

However, distributed systems – i.e., systems whose components might span across different machines and networks - have certain advantages, which include efficiency, scalability, and reliability, amongst others. In a distributed system, the autonomous nodes have local memory, and communicate and coordinate amongst each other through messages. Commonly, each node might have just a limited view of the system, or a specific function that it executes. In the latest version of the MENDO e-learning system for example, there are nodes called graders, whose job is simply to safely and securely grade a students' submission on a series of test cases (inputs and expected outputs). Because these nodes execute a single function, they might be automatically scaled up and down, as traffic increases or decreases.

A different usage of distributed systems is to share and synchronize content and courses among users, institutions or organizations [6], which leads to better content, network robustness, scalability, and improved privacy protections.

C. Privacy

Data privacy is important to both institutions and users. Recent regulations, such as, for example, GDPR, focus on empowering users in managing and protecting their data by giving them an option to intervene/control how and when their data needs to be removed, transparency (how data is being used), and minimization (i.e. systems should not ask for more personal information than is necessary for the purpose of the service) [7].

For e-learning, privacy norms also require that personal student information is only used for educational purposes, and that a student profile is only used in that context. Most software organizations also adhere to common security practices and follow guidelines and reports like OWASP Top 10 – a list of the most critical security findings and recommendations. Another common approach to tackling and limiting problems is to require and store only the information that is crucial for a system to provide the service for which it was designed – for example, if a student name or address is not required for an e-learning system to function correctly, then a simple username or student code can be used instead.

D. Item response theory

Item response theory is a paradigm of testing that focuses on items – as opposed to tests. IRT is concerned with the development of test items, where each item is not assumed to be equally difficult to solve. It is often considered as a superior alternative to classical test theory – for example, it is used in important tests like Graduate Management Admission Test (GMAT), Graduate Record Examination (GRE), Scholastic Aptitude Test (SAT), and others [8]. Often, for simplicity, we assume items to correspond to multiple-choice questions – but IRT can also work with scales and other complex scores.

For common tests used in schools, Item Response Theory leads to more accurate results with a smaller number of questions. For example, in Classical Test Theory, we can consider that the observed total score is based on the person true score (T) and a simple error (E), which is everything that is unrelated to the value we are measuring. On the other hand, we can consider IRT as a theory that allows us to create adaptive tests, and then to use the student's responses to improve the pool of test items. The standard mathematical model is based on the cumulative form of the logistic function – and based on the number of parameters, we can talk about a one parameter (1PL), two parameter (2PL) or three parameter (3PL) logistic model.

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Here, for the three parameter model, we outline that the probability of a correct response to an item i , depends on the item difficulty b_i (also called location), discrimination (a_i), and chance or pseudo-guessing (c_i). In the 2PL model, we assume that the data has no guessing – so it's appropriate for items where guessing is unlikely (the student has to enter an "answer", or a multiple-choice question which has many options to select from). Likewise, the 1PL model assumes that all items are similar in terms of discrimination.

Using this general idea, we can start with an estimated student ability (center, based on previous test results, or ask the user about their opinion), calculate which test item from our pool of items would be best and give that item as the next question/task for the student to solve. After the student completes the first test item, we update their estimated ability based on all previous responses, calculate the next best test item and repeat the procedure. This process continues until 1) the student completes a fixed number of questions, or 2) alternatively, the test might be limited in time.

In other words, when a student starts taking a test, an item of average difficulty can be presented to them. If they answer it correctly, they will get a more difficult item – or, if they answer incorrectly, they will get an easier item. This is the basic idea of Computerized Adaptive Testing (CAT), and many such systems rely on IRT.

One of the main goals of Item Response Theory is to update or improve the measurement accuracy and reliability of tests, either dynamically or using previously collected knowledge about student responses to certain test items. Comparably, with Classical Test Theory, reliability is much harder to estimate, because that would require knowledge about the student's true scores (minus errors).

In Item Response Theory, several assumptions are made in order to simplify the logic: 1) that a single trait, attribute or

ability decides the item responses and that it can be measured on a scale, 2) the local independence of items, and 3) that the response can be modeled by a mathematical item response function.

Some disadvantages of Item Response Theory are that it's relatively harder to compute and explain by hand (in comparison to scoring with classical test theory, which doesn't require any complex formulas), so therefore IRT tests are almost exclusively conducted with the help of computers. Additionally, IRT tests require a much larger database of test items, and a much larger number of students in order to obtain stable and valid results (the model and values improve over time).

Because of their ability to adapt to students, IRT is useful for developing and improving intelligent e-learning (tutoring) systems. Often, in e-learning systems, instructions are combined with assessments in order to increase retention in students, and to continuously evaluate and improve student models. For example, test items might be tagged/linked to certain instructions and lectures, or there might be a separate agent in the system which provides learning recommendations based on the IRT test scores (as shown in [9], where artificial neural networks are used for that purpose). The goal is for the e-learning system to guide the student to the correct instructions, and to adapt to their knowledge and skills – similarly to a human instructor.

III. ARCHITECTURE AND FEATURES

We have developed a distributed e-learning system based on item response theory, which allows students to learn different subjects and signup with either their name, email or – for privacy reasons - a username of their choosing (the choice can be made in consultation with their teachers). The system is composed of one main server component, developed with Node.js (using Express as the web framework) and utilizing a PostgreSQL relational database management system, where the main data is located (such as, for example test items), as well as gathered (anonymous) information from students utilizing the system. As a content management system (allowing both creation and easy editing) we decided to use WordPress. This content is then loaded in the main system by detecting changes through a sitemap.

For privacy reasons, and in order to make sure the system can be used even when schools have connectivity issues, each organization or school can run a separate web agent on their premises, which students connect to. These separate agents download the latest test items at configurable regular intervals (for example, daily), and can (optionally) send anonymized data to the main server component in order to improve the test items. The agents are also developed with Node.js (Express), but they utilize an SQLite database management system so that installation is easy. Both the main server and the web agents can be run inside Docker containers, and that is the preferred way of managing and running them.

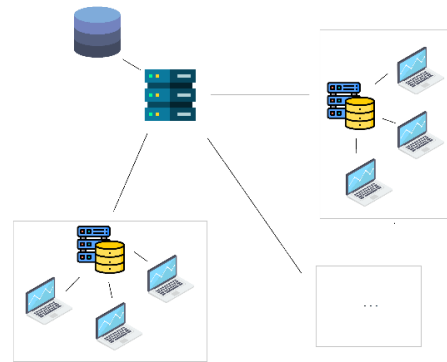


Fig. 1. High-level view of the system architecture

The client side was developed with Bootstrap and jQuery. We have made this decision based on our previous experience running other e-learning systems in Macedonia, and our gathered data showing that some operating systems and browsers used in schools are extremely outdated, so special care must be taken in order to make sure that students can use these systems both from home and from school.

Local storage is used to store a token that can be used to authorize and authenticate users in the system, and no tracking cookies or outside scripts are ever loaded when a student opens the web application. Both the main server system and the web agents use extensive caching, and we have tested and verified that the system can sustain more than a thousand requests per second – which can correspond to tens of thousands of active users at the same time for one agent (this number can be even larger, but it's dependent on the test items themselves, and the average time that users take to solve them – in addition to the server hardware). In our tests with real students, for which data is presented in the next section, we ran the main server system on Hetzner Cloud in one virtual private server, and the resource usage was minimal. This is one of the advantages of the event-driven and asynchronous nature of Node.js, making it suitable for real-time web applications which can sustain a very large number of active users at the same time.

The system uses AJAX requests in order to simplify the user interface and guarantee the best experience for the students. In the first version of the system, we only support one language (Macedonian), but both WordPress (used as a content management system) and the web server can easily be extended to support more languages – the main workload is actually to translate the entire database of test items.

Based on the architecture described above, and using item response theory, the system shows students test items (most of which are multiple-choice questions). After a student answers a question, (regardless if their answer was correct or not) the system shows them the correct answer and brief explanation of the idea, theory and solution (We have tried to keep this content brief, under 3 paragraphs, and to include images whenever possible to make learning more fun and engaging). Afterwards, based on the student response to the test item (correct or not), the system will offer them an easier or harder question. Because answers are saved in the agent database, the student can take breaks and use the system on different days if necessary – this is clearly one of the use cases for the system – i.e., for it to be used as an intelligent tutoring system. Before offering the student another question – the

system might decide to go with a question that might be slightly easier or harder than the optimal question to ask at that point, in order to guarantee that two students don't get the same list of questions if they make exactly the same answers (for example, imagine two students sitting next to each other, getting the same first question and making the same answer – to prevent cheating or gaming that system, in this case, our system might give a different question – which is based on slight randomization in the algorithm).

Note: It should be pointed out that in our tests and gathered data, the agents weren't running in schools, but on different Docker containers in one Virtual Private Server – with several subdomains pointing to a proxy running on the server, which then routed traffic to the correct agent/container. This was done in order to simulate a real working environment, but without asking schools and teachers to actually setup web servers and networks in their schools, as they might not be able to or have the necessary time or skills. Therefore, the system can be run either in schools, or on other servers as required.

IV. DATA AND SURVEYS

A group of 236 students participated in this research, 107 of them were part of the control group, and 129 tested the newly created e-learning system. Originally, 310 students signed up for participation, split evenly and randomly into the two groups, but some of them didn't complete all the necessary steps and surveys (outlined below) to be taken into consideration. Considering this was a weeklong experiment and didn't influence the students' grades (they were simply offered a certificate for participating in the experiment), it was expected that some of them wouldn't complete everything that was planned. The difference between the students from the control group completing the experiment (69%) versus the students using the system completing the experiment (83%) might indicate that the students found the e-learning system as offering real value for them and which will help them learn faster, but it might also be explained by temporary interest in something that is new to them.

Both the students from the control group and the students using the system were offered a web application with tests and study materials related to mathematics (due to the difficulty associated with creating a large set of test items and content related to their solutions, we decided to initially focus on one subject). The architecture and features of the newly created system were described in the previous section. For the control group, apart from test questions, the students had both pdf versions of learning materials and web pages containing the same lectures (as embedded pdf viewers) that they might reference while learning (to simulate standard learning from home). They were able to easily jump to any material or chapter after working on any question/task, or simply continue with the next one. Note that this is radically different than what the newly created system does – which automatically guides the students by choosing the tasks most appropriate for them, and then explaining the solution and brief theory behind it (all of which is previously created).

After the week-long experiment, students were asked if they agree with the premise that the web application that they used would be useful as an addition to traditional school learning. The results clearly indicate that students that used the newly created system showed much higher enthusiasm

and the scores in that group were much better compared to the control group.

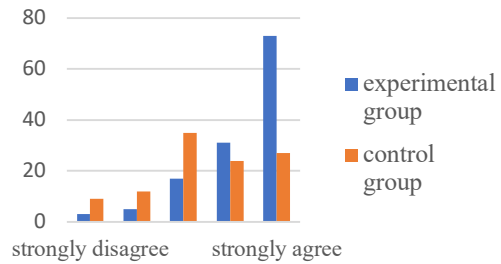


Fig. 2. Students' opinion about the willingness to continue using the web applications

Similar results were obtained when students were asked if they would like to use the system with other subjects besides mathematics, with the total scores being 4.132 for the group using the new system (if we consider 1 to indicate strong disagreement, and 5 strong agreement), compared to 3.318 for the control group using a simple web application with a classical testing functionality and static content.

Google Forms was used to conduct the survey, and students weren't asked for any personal information. Additionally, at the start of the form there was a message indicating that the survey is conducted anonymously.

Next, by analyzing the final data, we have noticed that students that used the newly created system spent 31% more time studying, which is a very significant difference. It should be noted that this estimate is based on the timestamps of submitted answers to test items, because we intentionally didn't measure other data and didn't track other student activities like the amount of time they stay on a certain web page, mouse movements, scrolling or clicks.

On the final day, we sent a test with 10 questions to both groups, and asked them to complete it in 30 minutes by themselves. These questions were actually taken from a real mathematics test, so that we don't pick questions ourselves and (possibly) introduce bias in the results. The experiment group showed better results than the control group, with the averages being 47 points and 41 points respectively (i.e. the experiment group had around 15% better scores). The maximum number of points was 100. Note that the student scores are positively affected by the fact that all tasks were converted to multiple-choice questions, and negatively affected by the fact that these tests didn't really influence the students or their grades in any way, so it is expected that they prepare much more for real tests later on.

Finally, with regards to the system itself, no significant downtime or issues were discovered during the testing period. UptimeRobot was used for monitoring, showing 100% uptime and expected/low response times. Similarly, the tracked cpu, memory and disk utilization was low – as expected. We also verified the communication between the web agents and the main server, and deliberately simulated communication errors (both by changing the api key, and blocking traffic), and all of them showed no errors or downtime. As soon as we reverted the api key and resumed traffic, the next scheduled data synchronization completed successfully.

V. FUTURE WORK

In the future, we plan to test this system with other subjects as well – we are already in the process of testing students in physics (where we also have the ability to add interactivity, and make a system that is much more interactive compared to reading a book or pdf). Similarly, we have plans for preparing materials for other subjects and classes in the near future. Although the system is designed in such a way that it can be utilized easily no matter the content - preparing a series of questions and solutions, as well as making the initial assessments of difficulty and discrimination, is extremely time consuming.

Additionally, the system currently only supports one language, and we plan to add support for more languages in the near future. Plans for future research involve adding more interactivity, notifications and viewing how they impact students, an ability for them to share their progress and possibly earn badges, and more.

VI. CONCLUSION

E-learning systems enable teaching with the help of electronic technology and resources, and are commonly used in combination with traditional methods of learning. There are several well-known learning management systems, some of which are open-source and freely available. Some systems, called intelligent tutoring systems, aim to provide automated and adaptable feedback to users, without the help or intervention from a human teacher. Privacy is also a key factor when talking about any web-based systems, but this is particularly important with regards to children and their education.

In this paper, we presented a distributed e-learning system based on item response theory. The system uses WordPress for preparing and managing content, and then functions through a main server component, and several agents which can be run either in the cloud or directly in schools. Clients connect to the agents, which can function even when they don't have communication with the main server component, as they store data locally in a SQLite database, and only communicate with the main server to fetch the latest test items, and to submit anonymized data so that the system can improve. Our tests and surveys indicate a very positive reception among students, and willingness to use such a system in addition to their traditional learning methods.

REFERENCES

- [1] Jando, Emanuel, Achmad Nizar Hidayanto, Harjanto Prabowo, and Harco Leslie Hendric Spits Warnars. "Personalized E-Learning model: A systematic literature review." In 2017 International Conference on Information Management and Technology (ICIMTech), pp. 238-243. IEEE, 2017.
- [2] US Department of Education OoP, Evaluation, Policy Development. Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. Washington: Office of Planning, Evaluation and Policy Development, 2009.
- [3] Cavus, Nadire, and Teyang Zabadi. "A comparison of open source learning management systems." *Procedia-Social and Behavioral Sciences* 143, pp. 521-526, 2014.
- [4] Kostadinov, Bojan, Emil Stankov, Marija Mihova, and Marjan Gusev. "State Competitions in Informatics and the Supporting Online Learning and Contest Management System with Collaboration and Personalization Features MENDO", 2013.
- [5] Bhattacharya, Suman, and Sayan Nath. "Intelligent e-learning systems: An educational paradigm shift.", 2016.
- [6] Van Steen, Maarten, and Andrew S. Tanenbaum. "A brief introduction to distributed systems." *Computing* 98, no. 10, pp 967-1009, 2016.
- [7] Zeide, Elana, and Helen Nissenbaum. "Learner privacy in MOOCs and virtual education." *Theory and Research in Education* 16, no. 3, pp. 280-307, 2018.
- [8] Chen, C. M., Liu, C. Y., & Chang, M. H. Personalized curriculum sequencing utilizing modified item response theory for web-based instruction. *Expert Systems with Applications*, 30, pp. 378–396, 2016.
- [9] Baylari, Ahmad, and Gh A. Montazer. "Design a personalized e-learning system based on item response theory and artificial neural network approach." *Expert Systems with Applications* 36, no. 4, pp. 8013-8021, 2009.