Forum for Education Studies 2025, 3(3), 1929.
https://doi.org/10.59400/fes1929

Perspective

Educational data mining in CALL assessment

Eunjeong Park

English Language Education, College of Education, Sunchon National University, Jeonnam 57922, South Korea; parkej@scnu.ac.kr

CITATION

Park E. Educational data mining in CALL assessment. Forum for Education Studies. 2025; 3(3): 1929. https://doi.org/10.59400/fes1929

ARTICLE INFO

Received: 11 December 2024
Revised: 22 February 2025
Accepted: 27 February 2025
Available online: 4 July 2025

Abstract: The deployment of data mining into computer-assisted language learning (CALL) assessment can help to transform language learning assessment and evaluation in a way it has never been. Advanced data analysis methods managed by machine learning and natural language processing can allow teachers and educators to view and analyze their language learning process data, making it possible to see many linguistic insights. The points of the theorem have a better vision that would be not only the learners' relationships with the digital education platforms but also the recognized value of the processed materials. Both teachers and learners can use text data in the analysis process to gain a macro view while the micro level of understanding is covered in many aspects. They help to determine and locate the levels of the language learning process that can act as a basis on which personalized feedback can be provided, with the students' different needs in mind. Furthermore, the data-driven CALL assessment as a means allows for the improvement of the language learning tools in terms of accuracy and equips instructors with tools to do their job better by improving their different teaching strategies. Using data mining in language assessment with technology in language learning facilitates teachers and educators to develop assessment training courses that can be used in various educational institutions worldwide and eventually leads students to an active learning process.

Keywords: CALL assessment; data-driven approach; technology; data mining; textual analysis

1. Data mining in CALL assessment

Because of the integration of technology, language education is considered one of the things that will probably receive a great change today. Language learning with the help of computer-assisted language learning (CALL) has established itself as a commanding pillar for the achievement of language competence [1]. CALL is defined as "an amorphous or unstructured discipline, constantly evolving in terms of pedagogy and technological advances in computer literacy and related literacies among both teachers and learners" [2] (p. 9). CALL investigates the role of digital technologies in mediating language learning and teaching [3]. Strengths of CALL include its capability of creating a situation or context in which learners working alone or in groups can interact with learning and reference materials. This interactivity puts the learning elements under learner control, allowing individual learners to approach the materials with their own agendas and actions [4]. CALL becomes an effective tool in language assessment with the benefits of relevancy, authenticity, and engagement [5]. Here, students have different types of interactive exercises, multimedia resources and personalized feedback at their disposal. In the context of the digital landscape, data mining is an emerging method allowing to investigate how effectively and how fast a learner progresses. In data mining, natural language processing (NLP) techniques are being implemented along with machine learning algorithms and data mining strategies to suppress the AI arbitrator who reviews the textual data from the activities of

language learning [6–8]. The case of assessment on CALL data mining has a prominent role in providing more efficient and tailored language education.

2. Opportunities

Personalized, real-time, feedback to learners is one of the crucial advantages of employing data mining in the CALL assessment. By reviewing learners' written answers, grammar mistakes and patterns of language application, data mining algorithms could contribute instant feedback on grammar usage, tempo of realizing vocabulary and overall language competence. The above-mentioned immediate feedback thus provides learners with the scope to identify their areas of improvement and tweak their language learning strategies. Moreover, data mining has challenged the definition of effectiveness, which means that learning materials can be modified according to learners' proficiency level, learning preference, and so on. Analyzing learners' interactions with digital learning materials through data mining will help identify their strong areas and weak spots, allowing educators to do fine-tuning of instructional content and activities in order to meet individual learning needs [6]. A one-to-one approach, with increased student involvement and passion, leads to better language skills acquisition. Moreover, as in the case with data mining, it shows the learners' progress during the period of time [1]. Language educators and professionals can examine students' written responses and students' language usage patterns across diverse educational activities. Based on the collected data, this machine learning helps to compile comprehensive reports and graphs illustrating learners' language proficiency growth trends. Therefore, this feedback enables teachers to supervise their students' development, identify the problems that persist and make sound didactic configurations based on that information. Data mining algorithms can also detect high-level language structures, rhetorical strategies, and learners' thought processes while producing language [8]. For example, using random forest theory allows teachers to analyze and excavate grammatical and syntactic features of the text, which can be used for further modeling and analysis of English grammar teaching. Other data mining algorithms are an English sentence classification method based on the decision tree method and an English grammar prediction model by using a large amount of English grammar teaching data for learning and classification [9–10]. This tiered approach brings more diagnosis on their proficiency in understanding the language and the cognitive activities engaged during the test, thus informing the revision of assessment tasks so that they can be more rigorous and authentic.

3. Challenges

Also, there are some challenges in regard to data mining that should be taken into account and solved before the CALL assessment [7]. First of all, interpreting test results can be difficult. In its essence, the purpose of the assessment is a right analysis and interpretation but contextual factors such as the context of learners' written responses, nuances of language use, cultural references and rhetorical conventions should be taken into account. Contextual ambiguity raises a number of issues related to the profundity of inferences and feedback. It arises when different groups of people in diverse cultures and languages may interpret and comprehend data based on their

unique cultural norms, linguistic constructs, or lived experiences. Therefore, the accuracy and reliability of the data inputs are central to data mining effectiveness in CALL assessment [7]. It will likely be difficult to balance the different language backgrounds, learning styles, and proficiency levels that learners have. Ensuring variations in this context are accounted for by algorithms presents a major hurdle towards achieving justice and accuracy in the results. In addition, data mining requires a robust technological infrastructure comprised of computational resources, software platforms, and data storage systems. Providing enough resources and skills to operate and maintain such systems might be a problem for institutions with limited IT capabilities [9]. Nonetheless, data mining can potentially boost CALL evaluation and eventually improve language learning effectiveness in any educational system.

4. Conclusion and implications

Data mining consists of a set of advantages in conducting evaluations for CALL. With intelligent upgrading, data mining reduces the assessment burden by using students' written works as feedback materials while the teachers have reduced workloads. AI feedback, enhanced by data mining and certain language-based skills such as word power, correct grammar and consistency, is meant for every learner to help them make improvements where they need it most. In addition to that, in applying text data to machine learning, a more adaptive learning environment is formed. Learners' mistakes and successes with the materials determine the materials that are given to them to get the best results. Discerning learning patterns influences the instructional material that is customized for the pupils, and the level of language competence that is assessed also sets them in the proper placement and monitors their progress. As to that, data mining usage in order to develop curricula designs is illuminated when educational materials are analyzed on a great scale. Secondly, it is for the purpose of classroom integrity and finally, it sheds light on some of the pedagogical models and language learning theories. As a result, data mining will have the great potential to raise the bar of assessment practices in CALL with efficiencies, effectiveness, and adaptability being the key points.

Variously, data mining is necessary in evaluating a learner's proficiency and progress in languages because of machine learning technologies. It is a great tool to offer introducing interactivity, personalization of experience, monitoring of the results and measuring the level of mastery of advanced language skills, which leads to setting up more effective and stimulating language learning environments. Data mining is turning out to be an inevitable trend in language learning, and text miners will definitely be leading the way in mastering the target languages by becoming more fluent and capable. By embedding these benefits of data mining, educational institutions can leverage the full potential of data mining to create adaptive and engaging language curricula. Cross-institutional collaborations could embody best practices and share innovations in curriculum development regarding data mining in CALL assessment that align with the demands of a digital and globalized world.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Yim S, Warschauer M. Web-based collaborative writing in L2 contexts: Methodological insights from text mining. Language Learning & Technology; 2017.
2. Beatty K. Teaching and researching computer-assisted language learning. Longman Pearson; 2010.
3. Yamazaki K, Thomas M. Computer-assisted language learning. In Oxford Bibliographies Online. Available online: https://www.oxfordbibliographies.com/display/document/obo-9780199756810/obo-9780199756810-0305.xml (accessed on 5 November 2024).
4. Espinosa VMG. "CALL" computer assisted language learning. Central Technical Institute Technological School; 2008.
5. Pinner RS. Teachers' attitudes to and motivations for using CALL in and around the language classroom. Procedia-Social and Behavioral Sciences. 2012; 34: 188-192. doi: 10.1016/j.sbspro.2012.02.037
6. Ahadi A, Singh A, Bower M, et al. Text Mining in Education—A Bibliometrics-Based Systematic Review. Education Sciences. 2022; 12(3): 210. doi: 10.3390/educsci12030210
7. Alhazmi HN. Text mining in online social networks: A systematic review. International Journal of Computer Science and Network Security. 2022; 22(3): 396-404. doi: 10.22937/IJCSNS.2022.22.3.50
8. Ferreira-Mello R, André M, Pinheiro A, et al. Text mining in education. WIREs Data Mining and Knowledge Discovery. 2019; 9(6). doi: 10.1002/widm.1332
9. Hassani H, Beneki C, Unger S, et al. Text Mining in Big Data Analytics. Big Data and Cognitive Computing. 2020; 4(1): 1. doi: 10.3390/bdcc4010001
10. Ai Q, Guo H. Intelligent Data Mining-Based Method for Efficient English Teaching and Cultural Analysis. International Journal of Mobile Computing and Multimedia Communications. 2022; 13(2): 1-14. doi: 10.4018/ijmcmc.293745