Design and Application of English Writing Training System Based on Web News Text Mining Technology

Yangxiameng Lu(✉)

Xi'an Medical University, Xi'an 710021, China

Abstract. Marking net, the largest online English composition marking platform in China, improves the speed of English writing feedback, and embodies the equal sharing of intelligence in learning. This paper focuses on the features of the new edition of the network, focusing on the technology of big data mining and learning analysis, discusses the application value of the visual graphics, charts and reports provided by the network, especially the classroom application value of error distribution and the role of guidance for accurate teaching and learning; points out that the ability of data analysis tests the multiple abilities of front-line teachers, which not only cultivates students' autonomous learning and learning abilityFinally, it briefly discusses the opportunities and challenges faced by intelligent technology itself.

Keywords: Correction network · Big data · Learning analysis · Error distribution · Accuracy

The Internet plus era is the era of rapid technological upgrading, and is the era of constantly updated ideas. Intelligent learning and intelligent teaching have gone deep into the behavior of teachers and students. However, the massive data provided by artificial intelligence is a challenge for teachers and students. "Data intelligence" provides technical support for accurate teaching in classroom, "teaching intelligence" will test teachers' decision-making and multiple action ability [1]."Data intelligence" also tests students' autonomous learning ability, constructs personalized adaptive learning habits, and conducts evidence-based intelligent learning. Customized teaching and learning based on big data analysis.

Teaching and learning will be the new mode and new power of learning in the future. This study specifically analyzes the application value of the new version of Junku correction network data in class and after class: through the bar chart, curve chart and word error report provided by the new version of correction network platform, the micro research on the two directions of data mining and learning analysis technology application of the new version of correction network is carried out, and it points out that only the integration of human and computer can implement effective correction [2].

An overview of the visualization and quantification of data in the latest edition of the online English composition and translation correction network, which is supported by corpus and cloud computing technology, is an online intelligent English composition

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 391, pp. 67–75, 2021.
https://doi.org/10.1007/978-3-030-87900-6_9

and translation correction service platform. At present, it is the largest English online platform in China, which is used by teachers and students in more than 5000 schools. The automatic online correction service greatly shortens the time for College English teachers to correct their compositions. It is no longer a headache for students to correct their spelling and grammatical errors. It provides a strong technical support for teachers' intensive speaking in class and accurate tutoring after class. Students don't have to wait for delayed feedback from teachers. The score and comment of the composition will be given in time for reference.

And comment by sentence, so that every student using this platform can enjoy the dividend brought by technological innovation, bathe in the sunshine of educational technology intelligence, and enjoy the equal rights of everyone.

Junku correction network is developed by Beijing ciwang Technology Co., Ltd., which can present the data of students' writing behavior in many dimensions. The data of the new edition of the network not only includes the visualization of scientific calculation, but also includes the content of information visualization and knowledge visualization. The abstract data is displayed on the screen in the form of graph (pie chart, histogram, curve) and other intuitive ways, realizing the goal of data visualization of different levels of evaluation [3]. This kind of data mining can effectively help teachers to carry out technical analysis, dynamically understand and master the actual learning situation of students, formulate teaching plans integrating offline, online and classroom based on the characteristics of students and the content of teaching materials, carry out data-based learning analysis, and implement accurate teaching. The figures and tables below show the data of the first composition of the new semester in March 2018. The students of No. 1070070 composition does plastic surrymake women truly beautiful? Are non English Majors of grade 2016 (mechanic 1607–1609 and special education 1601–1602). The first layer of the data (teacher interface) includes eleven dimensions: data overview, student performance, error distribution, browsing composition, similarity statistics, word frequency, collocation, graded vocabulary, data comparison, dimension analysis and retrieval. Click each dimension and the drop-down item contains different subitems. In the composition preview screen of the marking website, click the "more" option and select "commonality analysis" to display the above 11 dimensions. Next, according to the classroom evaluation, after class guidance, academic research these three perspectives for the new version of the correction network data mining analysis.

1 Classroom Evaluation

The five dimensions of "data overview, student performance, error distribution, browsing composition, similarity statistics" belong to macro data: including the information of the old version, such as submission, score, number of words, similarity, modification times. Next, analyze the classroom application value of these five dimensions."Data overview" provides three charts, the information has two parts, one is the chart, the other is the text: "submission statistics" histogram: including on-time submission (159), make-up (0), manual reading (0), similar (more than 40%) and text description "you received 159 compositions, a total of 28660 words, 1669 sentences, the longest sentence 75 words, the shortest sentence 3 words.

Score distribution curve: set score 60 as the starting point, 25 people as a group, divided into 5 segments. The information in Fig. 1 is as follows: 1 person scored less than 60, 8 persons scored 60–69, 50 persons scored 70–79, 98 persons scored 80–90, and 2 persons scored 90–100. The highest score of writing is 90.7 (Chen) and the lowest is 46.3 (MU). The average score was 80.6.

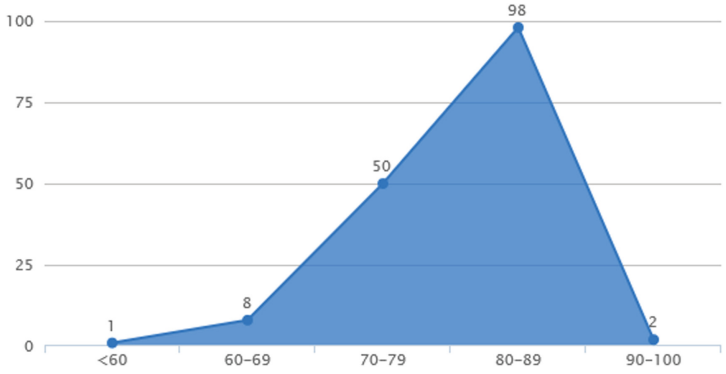

Fig. 1. Distribution of scores

Error statistics histogram: including the overall error text description, 399 errors were found in students' compositions. Among them, spelling and sentence composition are the most common mistakes made by students. Second, error statistics histogram. For example, 1070070 composition error distribution data histogram, according to the order from high to low, there are 17 sub items. The website visualized the general situation of this error release with a bar chart. The highest error was spelling error, with the data of 68. The lowest error was modal verb error, with the data of 1. (Fig. 2). There are many errors, which reflect the complexity of students' mistakes. It is related to students' personal English level and their attitude. If there are 68 spelling mistakes, students can correct them as long as they read carefully and follow the "warning" prompts. The data tells us that students need to correct their attitude to modify. They often type wrong words. Maybe they only pay attention to the meaning of Chinese when reciting words. If they memorize wrong words, they may make mistakes in typing. What's more, freshmen don't know much about English typing, resulting in wrong punctuation. Sentence structure, subject predicate consistency and misuse of part of speech are common mistakes made by Chinese students, which also reflects the challenges brought to Chinese students by different English and Chinese languages.

"Student performance" page includes expanding information, ranking (top 10), modifying ranking (top 10); using four colors to evaluate each student's individual situation (green stands for excellent, blue for good, yellow for general, red for poor), teachers can directly understand the overall performance of the students' composition by looking at the colors. Visual color data is convenient for teachers to analyze each student, and can also compare and analyze students. It is convenient for teachers to understand different classes of students as a whole by displaying visual information in intuitive colors. It

is easier to identify students' learning attitude, good performance and general performance by color difference, and these comprehensive information can be understood in an instant. Visual color data reflects that the correction network is actively deepening the artificial intelligence technology, reflecting the development direction of educational technology to a certain extent, and constructing concise, time-saving and large amount of information educational big data."Browse composition" page: including the composition score statistics, such as.

"More than 90 points (2 persons), 80–89 points (98 persons), 70–79 points (50 persons), 60–69 points (8 persons), and less than 60 points (1 person)". These data are the text expression in Fig. 2, showing each student's name, grade and modification times. You can click "comment by sentence" to view each student's composition.

Fig. 2. Error statistics

"Similarity statistics" page: a general table shows the repetition rate of all 159 people, and the information is displayed in the order of repetition rate from high to low, such as "more than 40% (5 people), 20–40% (9 people), 10–20% (20 people), and less than 10% (125 people)". For example, in No. 1070070 composition, Li's similarity is 79%, ranking first; 14 students have more than 20%, so they must rewrite it. There are serious problems in students' learning attitude, so we must criticize education. Objective and fair data give the criticism strength, which plays a very good warning role for students' academic norms and integrity in the future. The comprehensive information provided by these five dimensions is in the form of both graph and word. Data overview and student performance are graph data, error distribution includes graph and word information, which can be easily used in classroom composition evaluation. The overall performance of the students in each class can be objectively and accurately displayed in the classroom, explained at the same time, reasonably educated students, guided students to make good use of "comment by sentence", and strive to be honest and trustworthy students.

2 After Class Guidance

The second level data of the new version of the network is the refinement of the first level data, including different numbers of sub data, which provides macro and micro data for

teachers to comprehensively and accurately understand students. The error statistics chart (Fig. 2) and the word version of error distribution can be used for classroom analysis and individual student guidance. The data of this dimension makes clear the weak points of students' writing and provides good technical support for accurate after class guidance. At present, students' self adaptation, individual guidance strategies, error diagnosis and so on reflect the development direction of educational technology to a certain extent. Data mining and learning analysis is a hot topic in the field of international educational technology, and the domestic attention is also rising. With the combination of artificial intelligence technology, the network constructs its own intelligent data system. From a macro perspective, the "error distribution" in error diagnosis is the most characteristic. For teachers and students, the most practical type of data is "error release". The website provides a bar chart (as shown in Fig. 1 above). The data shows the distribution of students' writing errors from an overall perspective. The word version of the error evaluation report specifically shows each student's composition errors, points out the types of errors, and gives suggestions for revision. These wrong descriptions provide indispensable technical support for improving the accuracy of classroom teaching, one-to-one counseling after class, and the improvement of students' own ability. The combination of micro word and macro histogram provides reliable and evidential technical support for precision teaching. For example, the "growth track" (Fig. 2) records the historical version submitted by each student. Click on a specific student's growth track, the teacher can see the student's every revision, score change, submission time change, and can accurately guide the students to correct the composition, instead of 2 s, 3 s to submit, to see the score change. This reflects the students' speculative psychology. They want to get high marks, but they are not willing to spend more time revising. For students with too many modifications, viewing the historical version can accurately guide students and give pertinent guidance to their progress. For speculative students, criticize impetuous behavior and guide reasonable revision.

3 Academic Research

The last six dimensions "word frequency, collocation, graded vocabulary, data comparison, dimension analysis and retrieval" of the new edition of the website also include different numbers of sub items. For example, "data comparison" includes five sub items: "word statistical comparison, part of speech distribution comparison, collocation statistical comparison, sentence length distribution comparison and verb frequency comparison". For example, the word statistics comparison shows the comparison between a student and all the students. You can also compare the student with any one of the remaining 158 students in the same class or different classes. There are two kinds of data: histogram and statistical table. The reason why the data of six sub items are classified as micro level is that class explanation is limited by class hours and cannot be refined to this depth, but it is meaningful for deep academic research. From color histogram, curve chart, pie chart data to text description, it shows that the program design of the correction network is more refined and intelligent. As long as we users can understand the value of data, carry out data mining, and reasonably use these charts and tables for learning and analysis after class, we can overcome the disadvantages of poor adaptability and interactivity of data.

Second, the guiding role of data mining and learning analysis technology in teaching and learning.

With the advent of big data era and smart education mode, education reform has posed new challenges to front-line teachers. The teaching and learning mode of activities and tasks is no longer a single input, but the development and implementation of learning activities centered on students' autonomous learning. Modern education activities pay more attention to accuracy, autonomy, personalization and diversification. For teachers, only when they have higher ability of educational technology, actively participate in educational technology training, constantly explore in practice and constantly update ideas, can they not fall behind, better realize the role transformation, become the organizer of teaching and learning resources, the designer of process and the leader of behavior, and meet the requirements of important standards of teachers' professional quality. Since the report of enhancing teaching and learning through educational data mining and learning Analytics: an issue brief was released by the US Department of education in October 2012 [4], the academia has paid more and more attention to data mining and learning analysis technology. As the two directions of educational "big data" analysis and application, it is a late start and fast development discipline, technology empowerment learningLearning through technology [5] has become a new research path in the field of educational information technology and a new path to change learning. Sismesns and LAK (International Conference on learning analytics and knowledge) explain the concept of learning analysis from different aspects in the horizon report 2011 of NMC. Among them, the purpose of learning analysis has two prominent features: understanding the environment, optimizing the learning environment, measuring, collecting, analyzing and reporting the data generated in this situation [6]. He Kekang, a domestic scholar, has more concreted the exact connotation of learning analysis technology: "learning analysis technology tool, through collecting, measuring, analyzing and reporting massive data generated in the field of Education (the process of learning and the process of teaching management), extracts the hidden, potentially valuable, process and behavior information related to 'teaching and learning' or 'teaching management'It is a kind of information, knowledge and mode, so as to provide intelligent assistant decision-making technology for teachers' teaching, students' learning and teaching management [7]. The essence of learning analysis technology is to "support the implementation of evidence-based accurate teaching and effective teaching, emphasize the generation of efficiency or benefit, emphasize the guidance of teaching activities according to scientific principles, and strengthen the scientificity of teaching".

(1) The role of teaching guidance for teachers

A large amount of data about students' writing behavior has been obtained and stored in the marking network. No matter its quantitative data or qualitative data have reached a higher degree of unity and sharing of data structure and data format of different systems, with a unified data format standard and information model. A reasonable data analysis model has a high reference value for classroom teaching and accurate management of students. The integration of intelligent decision-making and Intelligent Implementation of intelligent teaching is inseparable from the support of intelligent technology. For example, teachers dynamically manage students' writing behavior. During the writing

period, teachers can read and correct students' compositions submitted online at any time. When they find that they are off topic or have similar red prompts, they can read and verify them. If the facts are consistent, they can cancel the composition, or click rewrite, and then use QQ in class.

The group small window informs them to rewrite, realizing real-time monitoring. You can check who didn't submit, how many students submitted last, and who didn't complete the minimum modification times by submitting date or student performance data. In the 1070070 writing, there are 6 students in class 07–09 who did not submit and did not make up for it. These 6 students scored zero in this composition; 25 students who submitted their compositions on March 28, the deadline, encountered the same problem in the process of submission: they did not submit successfully for many times, and they left a message on QQ to explain the situation. With the complete information obtained from data mining in Fig. 1, La technology can effectively help teachers to investigate whether students' performance and attitude are positive, and investigate students' integrity through similarity ratio, which provides a basis for timely education of students. Based on the technical support of correcting network data, learning analysis completely changes the biggest embarrassment of traditional manual correcting students' compositions plagiarism, and provides a scientific basis for teachers to make an objective, comprehensive, real and rapid evaluation of students. College English class, a teacher corresponding to a number of students, simply can not meet the requirements of everyone's correction, second correction. It is a good helper for teachers and greatly improves the feedback efficiency.

(2) The role of learning analysis technology in students' personalized and adaptive guidance

There are terminal consumers and students. As the largest user of the network, students' consumer experience should also be studied. When students submit their compositions, scores, comments and comments are generated instantly. For students, regardless of off topic, the first grade means the level of English writing, because in the student interface "my composition" window, the left side is students' composition (number of words, submission times), and the right side has the following four sub items: score, ranking, color viewable (yellow words, green sentences, blue chapter structure, light blue content related) and comments. Through the QQ survey, the students' feedback of "comment by sentence" has the most guiding role, and they think it is more practical. They modify according to the prompts, especially for those students whose grammar is not very good."Comment by sentence" is convenient for students to analyze their own learning and find out the relevant problems. Through many revisions, they can accumulate vocabulary and sentence patterns, and learn expressions in line with English Morphology and syntax habits, so as to improve their self-learning goal and improve their writing and translation ability.

The value of error distribution in three word versions Error distribution is an important data provided by the correction network. The data of visual histogram provides scientific and objective technical support for classroom precision teaching. In another word version of the same data, it clearly presents specific errors, comments and modification opinions, which is convenient for self modification. The teacher can get the

student's name by clicking hide. Error distribution data is the most practical. Both the histogram and the display of specific wrong sentences provide convenient and reliable technical support for learning analysis. This kind of data provides a diagnostic basis for teachers' accurate classroom teaching and face-to-face accurate guidance, which is convenient for teachers to use the most appropriate teaching methods and teaching strategies for error analysis. The suggestions on spelling mistakes in Fig. 5 are correct. Start pay is also wrong. The correct one is start to pay or start paying; "our society are also had many change", which is a typical Chinese English sentence, "are" "Had" does not exist in English, "have" as a state verb, basically does not use in passive sentences; "change" lost the plural, the appropriate expression should be "many changes have taken place/happy in our society". It is a common mistake for many Chinese students to mistakenly regard Chinese topics as English subjects. It is a difficult point for students to learn English and a key point in classroom teaching. In the classroom composition evaluation, in addition to the above analysis, we should give more examples to explain the role of Chinese topics, the methods of English translation, and explain with Chinese sentences without subject, which is helpful to improve students' writing and translation ability. English SVC sentence pattern is also a kind of error prone type for Chinese students. In Chinese, the adjective directly follows the subject. If I am happy, he is happy, there is no need for the copula be; in English, there must be be be to form the structure of the copula, SVCIt is one of the five core sentence patterns in English. Students are too familiar with it and make mistakes, which reflects that there are also problems in our teaching. We need to improve the explanation methods, emphasize the differences between English and Chinese, classify SVC and explain it systematically once, so as to avoid students making low-level mistakes again. When there are such errors, we can modify them by ourselves to improve the awareness of language self correction.

4 Opportunities and Challenges of Intelligent Technology

With more and more abundant data and higher intelligent performance, the feedback information from teachers and students in the new version is more detailed than that in the old version. For students, sentence comments and timely feedback improve their sense of participation, experience their own shortcomings in grammar and vocabulary, improve their awareness of prevention and optimize the learning process. Through the comments and the proportion of four colors, we can also understand the personal related problems and their performance in the class, which provides objective technical support for students' self-evaluation, and is conducive to personalized and adaptive learning, so as to improve the ability of English language expression. The development prospect of marking network is good, and its use value is very high for good, middle and upper middle school students. For teachers, the network faces more challenges. Many scholars have conducted in-depth discussion on this aspect, and I will not repeat it here. I just want to highlight two points: (1) the reliability and validity of the network.

It has been greatly improved. In the composition "how to tell Chinese stories in English" with the same title in May 2018, every student who misunderstands and doesn't write "how" has two red words of "digression". I have verified that the accuracy rate has reached more than 95%. In 1608 class, 8 students (8/30) only wrote stories related

to Chinese idioms, such as the tortoise and rabbit race, Mencius mother moving, farmer and snake, etc., but did not write "how to tell", and manual reading is also off topic. (2) Batch.

It is necessary to improve the export of the error distribution of network change. At present, the export of a composition is not achieved by class, but the total export of all the students' errors of a teacher, 1070070 exported 38 pages of word text. College English teaching is generally large class, a teacher has more than 200 students, which is not conducive to teachers' feedback in class. Accurate feedback in class requires manual pasting one by one, which is too time-consuming. For intelligent technology, one instruction and one programming can save the trouble of front-line teachers and achieve efficient feedback: QQ class feedback and accurate classroom feedback. The construction of a new model of precision teaching in class is inseparable from the high development of intelligent technology.

5 Conclusion

Ten year development plan of education informatization (2011–2020). It is clearly pointed out that "the teaching mode of deep integration of information technology and teaching" is the direction of university reform. English composition correction network provides effective technical assistance for College English Teaching in large classes. It is a good helper for front-line teachers, making the dream of everyone's feedback and timely feedback come true. As a gradually improved intelligent technology, it has been widely used in practical teachingIt provides students with a flexible, personalized and adaptive information-based learning environment anytime and anywhere, and also provides certain technical support for front-line teachers' accurate classroom explanation and online and offline accurate guidance. Whether data mining and learning analysis technology can be deeply integrated with the classroom and how to integrate them will become an important mission of big data analysis in education, and also a key factor in building a lifelong learning system and a learning society.

References

1. Zhiting, Z., Hongchao, P.: The rise of smart learning ecosystem research. Chin. Audio Vis. Educ. **6**, 1–11 (2017)
2. Hongmei, L.: The value logic of "new" learning methods in the "Internet+" era. Chin Audio Vis. Educ. **6**, 102–107 (2017)
3. Bing, Y., Guoqing, L., Honggen, X., et al.: Research and implementation of data visualization in online learning system — taking sat platform as an example. Mod. Educ. Technol. **12**, 114–120 (2017)
4. Kekang, H.: New development of "learning analysis technology" in China. Audio Vis. Educ. Res. **7**, 5–13 (2016)
5. Kelin, L., Zhimin, X.: Technological change learning from the perspective of empowerment: a review of the "learning" part of the national educational technology program 2017 update. Mod. Educ. Technol. **3**, 26–32 (2018)
6. Yonghe, W., Dan, C., Xiaoling, M., et al.: Learning analysis: a new wave of educational informatization. J. Dist. Educ. **4**, 11–19 (2013)
7. Ronghui, G., Gaoda., H.: Research prospects of learning analysis and foreign language teaching in the era of big data. Audio Vis. Foreign Lang. Teach. **3**, 40–45 (2016)