Electronic Dictionary Research Corpus

Wei Sun*
Southampton International College, Dalian Polytechnic University, Dalian City, P. R. China

*Corresponding author e-mail: sunwei@dlpu.edu.cn

Abstract. The development of computer and Internet technology, and the emergence of new technologies such as electronic dictionaries and corpora have made it more convenient, accurate and reliable for students to learn to access materials. Students must master these new technologies in their study and life. The purpose of this article is to focus on the research and analysis of the electronic dictionary corpus, and to understand the influence of the electronic dictionary corpus on students' study and life. The inquiry method used in this article is to conduct a questionnaire survey on students using the electronic dictionary corpus and then analyze the data obtained through big data technology. At the same time, this article still uses the simulation method to predict the future development prospects of the electronic dictionary corpus with the computer simulation experiment. The electronic dictionary corpus uses a variety of algorithms to query massive databases, and finally can display the data we need in front of us at the fastest speed. The dynamic BP network is used to predict the generalized predictive control prediction error. The dynamic BP network the weight can be adjusted online, which changes the situation that the traditional BP network performs error correction once the network weight is fixed after training. The experimental results show that this article conducted interviews and questionnaire surveys with experimental students and found that more than 90% of the students believe that the electronic dictionary corpus has a positive impact on students' learning and development.

Keywords: Electronic Dictionary Corpus, Computer Simulation Experiment, PID Network

1. Introduction
As we all know, corpora is not a new thing in our country. Since Chinese scholars introduced it into the country, the research and application of it in various fields have been in full swing, and academic attention has also increased exponentially. The establishment of various types of corpora has also It fully shows that this digital tool has been accepted by various disciplines, and the electronic dictionary formed by the use of corpus for lexicography is its fast-developing branch [1, 2]. The construction of fast-growing electronic dictionaries also uses the computer platform to continuously expand the extension of the original reference books in the traditional sense, and the continuous innovation of query methods and concepts has brought breakthroughs in the use of reference books [3, 4].
Scholars at home and abroad have different views on this. In the traditional sense, foreign scholar Jun S believes that when traditional dictionaries encounter strong competitors like electronic dictionaries, their survival will be difficult. Therefore, whether it is newspapers or TV when encountering this problem, they have launched their own online editions one after another, and the more traditional reference books chose to integrate in an updated way when facing the opponent of the computer [5]. It not only formed its own electronic version, but also made a lot of effort on the electronic version [6]. In order to explain this problem better, the domestic scholar Huang H, first of all, the promotion of electronic dictionaries is to a considerable extent representative of the digitalization of the entire reference book [7]. Secondly, it is its scale. CNKI has collected more than 2,000 reference books. To a certain extent, it has covered all the reference books that can be covered in the current market [8]. And the collection of reference books is still in progress. Striving for all is also the goal pursued by CN2KL. The third point is because people have made breakthroughs while making electronic reference books. They converted paper reference books into electronic versions at the same time. A knowledge network is formed in its own database, which is a breakthrough and bright spot of CNKI [9, 10].

Based on the questionnaire survey, this paper conducts a comparison experiment and analyzes the advantages and disadvantages of students using electronic dictionary corpus through random sampling and dynamic BP network error correction. Statisticians have proved that the accuracy of sampling analysis will greatly improve with the increase of sampling randomness, but it has little relationship with the increase of sample size. However, in practice, random sampling also has unseen problems, which are often caused by the sampling design not suitable for the structure of the data and noise. Sampling based on known probabilities, no matter whether the probabilities are equal at the time of sampling, whether it is a one-time sampling or sequential sampling, is called random sampling. The premise of applying random sampling is that the probability of the sampling survey design plan can be explained through decomposition the model is used to estimate the overall parameters of interest on a large scale, even if in terms of design, some observed individuals have a greater probability of sampling than others. As for those sampling deviations caused by unequal weights, other models such as setting sampling weights or using appropriate regression models can be used to eliminate them during estimation. Moreover, deviations can allow people to use multiple algorithms to query massive amounts of electronic dictionary corpus. The database can finally display the data we need in front of us at the fastest speed, which greatly improves our learning efficiency.

2. Electronic Dictionary Research Corpus

2.1 Principles of Random Sampling Survey Method for the Use of Electronic Dictionary Corpus

Many of the data features displayed in large-scale data sets are non-random. Some of these features may be "appearance", and some may be the "essence" of real things. I want to extract the real from various features. Essential characteristics still need to be selected randomly, because the essential characteristics of everything are displayed randomly in nature, which is an unchangeable natural law. For the existing large-scale data set, we can extract all the corresponding data according to the research problem, but these data are not necessarily the overall data of the research object, may be part of the infinite population, or may be the finite population in this case, the safest way is to treat this part of the data as a large-scale sample, and then use the bootstrap method to repeat sampling. The basic idea of the bootstrap method to get the closest overall eigenvalues is: First, set a sample X from the population distribution F. R(X,F) is a preset function of X and F, which can be the mean, variance, etc., and then, when the population distribution is unknown, Using the bootstrap sample x generated by repeated sampling with replacement to estimate the distribution characteristics of R(X,F). Bootstrap samples are multiple samplings with replacement on an existing large-scale data set. Generally, the number of extractions selected is not less than 1000. If the sample size is n, each observation in the data set will be the probability of drawing is 1/n. According to the bootstrapped sample, the sample
estimated value θ of the overall parameter Θ can be calculated. Assuming that the sampling is repeated M times, then the bootstrap estimates of the mean and standard error of the parameter Θ are:

\[ \Theta = \frac{1}{M} \sum_1^M \theta \]  

\[ se_\theta = \left[ \frac{1}{M-1} \sum_1^M (\theta - \Theta)^2 \right]^{1/2} \]  

The above bootstrap method only uses the asymptotic theory to obtain the specific data of the estimator, but we are more concerned about the distribution of the estimator, so that we can find the optimal estimator. For this problem, we can use the estimated overall distribution to regenerate a set of random variables, if it is a good enough estimate, then the relationship between X and F can be well reflected in the relationship with. Repeating the above steps several times will find multiple new estimates similar to θ, and then by analogy, find the most accurate estimate. Obviously, using this method to obtain estimates requires a lot of repeated calculations. For large-scale data sets containing thousands of pieces of data, the computational cost cannot be underestimated. Therefore, a Bag of Little Bootsprts method is proposed. On the basis of retaining the universality, statistical validity and theoretical characteristics of the bootstrap sampling method, this method improves the computational efficiency and the robustness of the evaluation estimator, and is very suitable for distributed parallel computing architecture.

2.2 Principles of PB Predictive Analysis Method Used in Electronic Dictionary Corpus

The error predicted by the dynamic BP network is \( y_e(k + j) \), which is used as the error compensation to correct the generalized predictor:

\[ y(k + j) = y_m(k + j) + y_e(k + j) \]  

\( y_m(k + j) \) is the predicted value of the traditional generalized prediction algorithm at time k. Introduce (3) formula:

\[ y(k + j) = G_j(z^{-1}) \Delta \mu(k + j - 1) + F_j(z^{-1}) y(k) + y_e(k + j) \]  

The optimal solution after error compensation is:

\[ \Delta U = (G^T G + \lambda I)^{-1} G^T (Y - F - Y_e) \]  

The dynamic BP network is used to predict the prediction error of generalized predictive control. The weight of the dynamic BP network can be adjusted online, which changes the situation that the traditional BP network performs error correction once the network weight is fixed after training.

3. Experimental Research on Electronic Dictionary Corpus

3.1 Experimental Data

This research will use the status quo of middle school students’ dependence on electronic dictionary corpus tables on middle school students’ electronic dictionary corpus and its impact on academic performance and mental health. Invested in the discussion of this issue, to provide a reference for in-depth study of middle school students’ mobile phone dependence behavior and education issues.

3.2 Experimental Process

The design of the questionnaire survey is divided into two types: student questionnaire and parent questionnaire. The subjects of the survey were selected randomly selected students from three schools and their parents. The survey question is the influence of the electronic dictionary corpus on student performance. The survey received 658 questionnaires from students and 63 questionnaires from passersby. First of all, this article randomly selected students conducted a questionnaire survey to
obtain the attitudes of these 658 students on the use of electronic dictionary corpus. Finally, conduct a questionnaire survey and compare experimental data.

4. Experimental Analysis of Electronic Dictionary Corpus

4.1 Views of Students and Parents on the Use of Electronic Dictionary Corpus

This article uses the questionnaire method to conduct a questionnaire survey on 658 randomly selected students. In order to obtain contemporary students and parents' understanding, like and recognition of electronic dictionary corpus in the new media environment, they can have a more true understanding of their views on the use of electronic dictionary corpus. The purpose of the first questionnaire survey is to understand the influence of students and parents on the use of electronic dictionary corpus in the use of electronic dictionaries, and the second questionnaire survey is to deal with changes before and after the use of electronic dictionary corpus. The survey results are shown in Table 1 and Figure 1.

Table 1. Views of Students and Parents on the Use of Electronic Dictionary Corpus

|                | Understand resource pool | Used resource library | Like resource pool | Think resource pool is very important |
|----------------|--------------------------|-----------------------|-------------------|---------------------------------------|
| Student        | 156                      | 106                   | 42                | 128                                   |
| Parent         | 108                      | 90                    | 36                | 78                                    |

Figure 1. Discussion on the Use of Electronic Dictionary Corpus By Students and Parents.

It can be seen from the survey data that most students know very little about the use of electronic dictionary corpus, do not know much, and rarely have contact with it. Therefore, only a few students express their appreciation for it. On the other hand, students cannot realize that the electronic dictionary corpus makes important decisions for future students’ learning and development, and integration into the new learning environment will inevitably affect the changes in students’ learning styles in the future. Traditional learning is boring, so everyone doesn't like the traditional learning method very much, and they don't realize that the electronic dictionary corpus will have a huge impact on the changes of future students' learning methods.

4.2 Changes of Parents and Students' Perception of Electronic Dictionary Corpus
In this article, parents and students are exposed to the use of electronic dictionary corpus for a one-month simulation. During the experiment, the students and parents participating in the experiment are surveyed every 5 days, and the changes in their perception of the use of electronic dictionary corpus are counted. This article visualizes the degree changes caused by the use of electronic dictionary corpus by parents and classmates, and performs curve fitting according to the mean value. As shown in Table 2, Figure 2

Table 2. Every Five Days, Parents and Students' Perception of the Use of Electronic Dictionary Corpus Changes

| Time      | After 5 days | After 10 days | After 15 days | After 20 days | After 25 days | After 30 days |
|-----------|--------------|---------------|---------------|---------------|---------------|---------------|
| Parent    | 50%          | 55%           | 63%           | 75%           | 81%           | 86%           |
| Student   | 50%          | 51%           | 54%           | 60%           | 63%           | 69%           |

Figure 2. Visualization of Changes in Parents and Students' Perception of the Use of Electronic Dictionary Corpus Every Five Days

The experimental results show that over 90% of the students who conducted interviews and questionnaire surveys on the experimental students believe that the electronic dictionary corpus has a positive effect on the learning and development of students.

5. Conclusions

This article explains that nowadays massive amounts of data have penetrated into the study and life of all walks of life. People should focus on how to use these data resources to provide more accurate information. Many of them will have various difficulties. Electronic dictionary corpus can be used as a way to solve some of the difficulties, such as saving data storage space; obtaining data in unknown fields; exploring the causal relationship between data; improving data quality and query speed. The
data itself is relatively large and small, more and less, and the evaluation of big data and sample surveys cannot be based on a single perspective. The electronic dictionary corpus can more accurately find the knowledge we need, which greatly saves our traditional data-searching world and improves our learning efficiency. Traditional learning is boring, so everyone doesn't like the traditional learning method very much, and they don't realize that the electronic dictionary corpus will have a huge impact on the changes of future students' learning methods.

References
[1] Shinar J, Steinberg D. Analysis of Optimal Evasive Maneuvers Based on a Linearized Two-Dimensional Kinematic Model. Journal of Aircraft, 2015, -1(8):546-554.
[2] Hwang E C . Web Based Smart Home Automation Control System Design. International Journal of Contents, 2015, 11(4):70-76.
[3] Raulji V D , Pujara H , Arambhadiya B , et al. Automation of Aditya vacuum control system based on CODAC Core System. Fusion Engineering & Design, 2016, 112(Nov.):910-914.
[4] Michal, Sobolewski, Norbert, et al. Fuzzy logic estimator implemented in observation-tracking device control. Aircraft Engineering and Aerospace Technology, 2015, 88(6):697-706.
[5] Jun S , Meixia Z , Zhengming L , et al. Simulation of Smith fuzzy PID temperature control in enzymatic detection of pesticide residues. International Journal of Agricultural and Biological Engineering, 2015, 8(1):50-56.
[6] Garrido P , Soares F , Moreira, António Paulo. [Lecture Notes in Electrical Engineering] CONTROLO 2016 Volume 402 || Smith Predictor Tuned Through Fuzzy Fractional PID Controller. 2017, 10.1007/978-3-319-43671-5(Chapter 7):71-80.
[7] Huang H , Zhang S , Yang Z , et al. Modified Smith fuzzy PID temperature control in an oil-replenishing device for deep-sea hydraulic system. Ocean Engineering, 2018, 149(FEB.1):14-22.
[8] Osinski C , Villar G , Da Costa G . Fuzzy PID Controller Design for LFC in Electric Power Systems. Latin America transactions, 2019, 17(01):147-154.
[9] Np A , K, B . High performance hybrid fuzzy PID controller for permanent magnet synchronous motor drive with minimum rule base. Journal of Vibration & Control, 2015, 21(1):181-194.
[10] Osinski C , Leandro G V , Oliveira G H D C . Fuzzy PID Controller Design for LFC in Electric Power Systems. IEEE Latin America Transactions, 2019, 17(1):147-154.