Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 4, pp. 2451–2462, DOI 10.12694/scpe.v25i4.2864

ANALYSIS OF LIFT-APRIORI-DP JOINT ALGORITHM-BASED DATA EXTRACTION IN BUSINESS ENGLISH ACHIEVEMENT IN COLLEGES AND UNIVERSITIES

HONGYING XIAO*

Abstract. This paper investigates the application of data mining based on a correlation-rule algorithm in business English performance in colleges and universities. The extracted correlation degree rules are screened by adopting three indexes of support degree confidence degree and lifting degree to measure the correlativity. Experimental validation is carried out on different sets of data sets, and the experimental results show the effectiveness of the Lift-Apriori-DP algorithm. Based on the improved Lift-Apriori-DP algorithm, it is applied to the analysis of students' performance. Taking the chapter test scores of students in business English courses in colleges and universities as an example, the student's achievements are extracted and analyzed, and the final appropriate parameter values are selected according to the evaluation of the extraction results under different privacy pre-estimation. The experimental results verify the effectiveness of the Lift-Apriori-DP algorithm in the analysis of student grades and evaluate the accuracy of the algorithm application. The results of this paper show that data mining based on the correlation degree rule algorithm has a wide range of applications in business English grades in colleges and universities, which can provide useful references for teaching and at the same time protect students' private information from being leaked. In addition, this paper also explores the evaluation class method of the mining results under different privacy pre-estimation, which provides a useful reference for the application of privacy-protecting relevance degree rule mining type algorithms.

Key words: lift-Apriori-DP; achievement big data; data mining; privacy preservation; law of association

1. Introduction. With the continuous progress of science and technology, massive data resources have been gathered in today's era [1-2], and the scale of data resources is still continuously expanding. How to explore the potential information behind the data, data mining technology [3-4] has emerged. The hidden information behind educational data can be more intuitively extracted through the use of data mining techniques to improve the quality of teaching and learning [5-6]. Among various data mining algorithms, the relevance rule mining algorithm [7-8] is one of the important branches, that has been widely used in various industries. Relevance rule mining algorithms aim at mining interesting relevance relationships in transaction databases to help people better utilize the knowledge behind the data, which is an indispensable guide for teaching and education, business decision analysis, evaluation and analysis of the healthcare industry, and the enhancement of higher education. A classic example is "80% of the customers who buy bread also buy milk". The most commonly used algorithm in the relevance rule extraction algorithm is the Apriori algorithm [9]. In practical applications, the L-Apriori-ll algorithm is relatively concise and clear, and can better discover the implicit rules between the data, for example, in higher education and other industries to analyze the student performance data for relevance degree rule mining, and early warning of student performance [10]. At the same time, with the development and practical application of data mining technology, the problem of privacy data leakage becomes more and more serious when mining the hidden information behind the data [11-12], Dalenius [13] firstly put forward the protection of data privacy for data security problem, and its main point is that during the period of querying the data, for any data user, whether it is legal or illegal user, can not get the information of the data. Dwork first proposed the concept of differential privacy [14]. Differential privacy has two outstanding advantages, on the one hand, it does not need to take into account the knowledge of the background possessed by the attacker; on the other hand, the model is based on rigorous mathematical calculations with quantitative evaluation class methods, and thus it is gradually being widely used [15].

In 2015, Li11 [16] and Gan [17] improved the efficiency of the algorithm by mapping the data to a relatively

*Faculty of humanity and law, Gannan University of Science and Technology, Ganzhou Jiangxi 341000, China (hongyingxhy@outlook.com)

low dimensional space and reducing the dimensionality by truncating the transactions, respectively, and Cheng [18] proposed a frequentness class-term group mining-type algorithm based on the preservation of differential degree privacy by transaction splitting. In 2017, Zeyu Shen et al [19] introduced a parallelized environment, which mainly performs parallel computation under the Hadoop framework, and the computational effectiveness of the algorithm is significantly improved when mining large-scale data. In 2018, Yihui Cui and Wei Song et al [20] investigated a multi-source data relevance rule mining algorithm under the protection of differential privacy, which performs noise perturbation through the Lap-Las mechanism that publishes data in the form of a tree of frequency class itemsets with high security. In 2019, Han et al [21] investigated an association degree rule mining-type algorithm for distributed databases with differential degree privacy protection by adding Laplacian noise to the counts of frequent 1-itemsets for each candidate itemset set. In 2019, Tingting Chen et al [22] proposed a heuristic algorithm for frequentness class itemset mining with differential degree privacy, which performs heuristic truncation of transactions to reduce the degree of global sensitivity, and the improved algorithm has a high degree of usability. In 2019, Chen Jiang et al [23] proposed the Trun-Super algorithm to address the low efficiency of long transaction mining in the frequentness class item mining-type algorithms oriented towards privacy, and privacy protection, which mainly reduces the efficiency of the data through the transaction data group set intelligence truncation for dimensionality reduction.

In the field of educational data extraction, the research on the correlation degree rule has achieved certain results. The research mainly lies in the analysis of teaching data through the L-Apriori-ll algorithm based on the L-Apriori-ll algorithm and the improvement of the L-Apriori-ll algorithm according to the problems in the practical application. Analyzing the extracted relevance degree rules provides guidance suggestions for future teaching. The following are the applications of association degree rule algorithms in the field of educational data mining. In 2005, Markellou et al [24] proposed a personalized learning framework for the problem of personalized online learning by using the L-Apriori-ll algorithm to find the group of frequency class items that can better recommend learning materials for users. In 2009, Huo Shuhun et al [25] extracted and analyzed the teaching degree evaluation data accumulated in higher education institutions, and analyzed the implied correlation relationship between the quality of teaching and the teachers' related information according to the correlation rule extraction algorithm, to motivate the teachers to better improve the quality of teaching [26]. 2010, Buldu et al [27] used the L-Aprori-ll algorithm to find the frequency class items of the group of students in a high school business class in Istanbul and proposed a personalized learning framework. Istanbul Higher Business High School students' achievements were mined and analyzed to help students make independent career choices by revealing data relationships between students' social activities, areas of interest, and unsuccessful courses. In 2011, Milicevic et al [28] proposed a personalized e-learning system capable of automatically adapting to the learner's interests, habits, and level of knowledge using the Apriori-All algorithm, by mining and analyzing data from the Istanbul Higher Business High School students' achievements and helped students to make independent career choices. -All algorithms to recommend learning programs adapted to learners by mining the frequency sequence in each learning habit to improve the quality of teaching and learning. In 2012, Xiao [29-32] analyzed the data using the Apriori algorithm for the stored teaching-degree evaluation data in distance education to derive the key evaluation data in the mining results that have an impact on the quality of teaching and learning, which in turn improves the teaching-degree evaluation quality. In 2018, SungSik et al [33], used L-Apriori-ll to extract and analyze students' math performance and problem-solving patterns and analyzed the results of the extraction to select the information that is valuable for improving students' performance. Cheng-Yong Wang [34] targeted the L-Apriori-ll algorithm running with low efficiency due to multiple scans of the D database, etc., and proposed an altered procedure for control.

In practice, the Apriori algorithm based on differential privacy preservation often mines a large number of invalid rules [35-36], mainly because of some problems in filtering segments based on support and confidence. To solve this problem, this paper introduces the lifting degree to improve the Apriori-DP algorithm. Boosting degree can measure whether a rule is better than a random guess, thus helping us to better filter out valid rules. In addition, to better protect students' private information, this paper adopts the Apriori algorithm based on different degrees of privacy protection for data mining. This algorithm can ensure that students' private information is not leaked, and at the same time, it can get relatively accurate results. To summarize, in the field of educational data mining, the research on relevant rule-based mining algorithms based on the

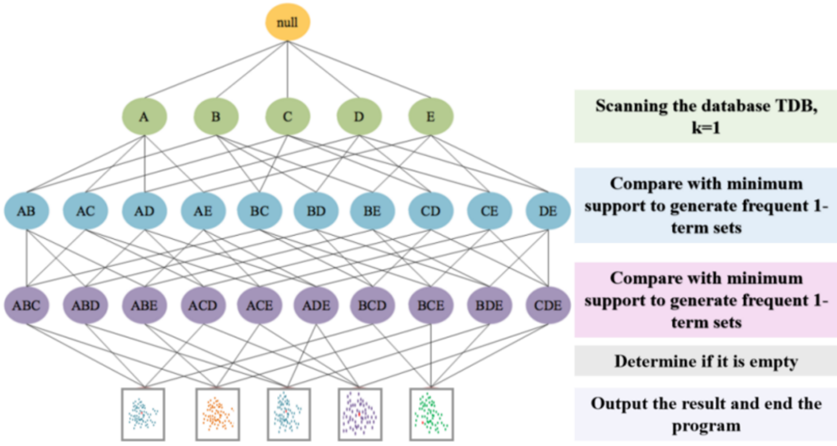

Fig. 2.1: A priori-based algorithmic flow

protection of differential degree privacy has practical application value. Therefore, in this paper, a data mining model based on the algorithm of relevance rules is developed for the achievement of business English, which can quickly realize the intelligent analysis of the achievement while guaranteeing privacy.

2. Lift-Apriori-DP algorithm.. This chapter first describes the problem that in the Apriori algorithm based on the protection of differential degree privacy, many invalid correlation degree rules are extracted when clips are screened according to the support degree and confidence degree. Secondly, the lifting degree is introduced to improve the Apriori-DP algorithm. Then the design of the Lift-Apriori-DP algorithm with the introduction of the lifting degree is introduced. The experimental analysis is carried out according to the improved algorithm, and the experimental results show the effectiveness of the improved Lift-Apriori-DP algorithm. Finally, the Lift-Apriori-DP algorithm is applied to the analysis of students' performance to analyze the effectiveness of the Lift-Apriori-DP algorithm in the analysis of student's performance and to assess the accuracy of the algorithm application.

2.1. Lift-Apriori-DP New Algorithm Flow and Logic.. The basic idea of the new joint Lift-Apriori-DP algorithm is divided into two main steps: one is to find the frequency class itemsets, and the other is to generate the strong-associative degree rules. First, the algorithm scans the discrete database and generates a frequency 1-item set L. Then, it uses L to generate candidate item sets A/B/C/D/E items, etc., accumulates the frequency of item sets of C2 and compares it with the minimum support, removes non-frequent item sets, and generates items such as AB, i.e., frequency 2-item sets. This process is looped until no more frequency k-item groups can be found. The entire database needs to be scanned once for each frequency class item group found. The specific process is shown in Figure 2.1.

The core of the Lift-Apriori-DP new joint algorithm is to connect and clip the frequency class item group lookup, mainly layer-by-layer search iteration, the algorithm is simpler and clearer, more suitable for transaction database mining analysis, and thus is widely used in various fields of relevance of the degree of the rule of mining analysis, pseudo-code algorithm description as shown in Table 2.1.

2.2. Differential Processes for Privacy Protection.. Differential degree privacy model for privacy protection provides a higher level of semantic security and, a greater possibility to protect data security from successful attacks, and thus is widely used in various aspects. The relevant definition of the differential degree privacy model in the model of this paper is as follows:

There are two data sets D and D', when the data records in D and D' satisfy equation 2.1, i.e., D and D'

Table 2.1: Process control code for the joint Lift-Apriori-DP algorithm

Lift-Apriori-DP New Joint Algorithm Code
Input: data set TDB, support level min_sup Output: frequency class item set
1.// Scanning TDB to get C
2.C, = find_candidate_1_itemsets (TDB); for Vt eTDB{
// Scanning the TDB for counts
3. C, = subset(C„ t); // generate C
4. for VcEC,;
5. c.count++;
6.7.8. l, = {ceC, Ic.count/t_num 2 min_sup)//C, clip generates L, for (k=2; L+O; k++){
9.C= apriori_gen(L): // Generate C from L self-connection
10. for Vt eTDB{
11.C, = subset(Ck, t); // get the subset containing the subset over C
12. for VceC.
13.c.count ++;
14.}
15.L={ceC, lc.count/t_num 2 min_sup)//from C, generate L
16.17. return L =ULk

differ by one data set, then they are called neighboring data sets.

$$|(D-D')\cup(D'-D)|=1 \tag{2.1}$$

The parameter ε is a positive real number and is a privacy-preserving pre-estimate. There is a randomized algorithm M, and PM is the set consisting of any outputs of the randomized algorithm M. For any two neighboring dataset sets D and D', as well as the set of outputs P, and any subset Sy of it, a randomized algorithm M is said to protect e-differential degree privacy if the effect of the randomized algorithm M on the neighboring dataset sets satisfies Equation 2.2.

$$|(D-D')\cup(D'-D)|=1P_r\left[M(D)\in S_M\right]\leq\exp(\varepsilon)\times\left[M\left(D'\right)\in S_M\right] \tag{2.2}$$

Figure 2.2 shows the output probabilities of the randomized algorithm M acting on D and D'. The choice of the size of the pre-estimated value for the degree of privacy is very important. The higher the degree of privacy protection required by the algorithm, the smaller the value of e is, the better, but the degree of usability of the algorithm's extraction results is relatively lower; relatively speaking, when c is larger, the degree of usability of the data extraction results is better, and the effect of privacy protection is relatively lower. In the practical application of the algorithm, the privacy pre-estimation limit boundary value should be selected reasonably according to the demand.

Let x-Laplace (μ, b), where the position parameter is μ and the scale parameter is b, and its probability density function is Equation 2.3:

$$f(x)=\frac{1}{2b}\exp\left(-\frac{(|x-\mu|)}{b}\right) \tag{2.3}$$

The distribution of the Laplace probability density function for different parameters (under input and output values) can be observed in Figure 2.3 below. By presenting the probability density distributions under different parameter points P, respectively, it can be found that the relationship between the input and the output is plotted as shown on the left side, and the simultaneous transformation leads to the Laplace-Las probability model shown on the bottom right. It can be seen that: in the position parameter, the same case, the scale parameter is the smaller b, the smaller ε, that is, the larger the noise introduced.

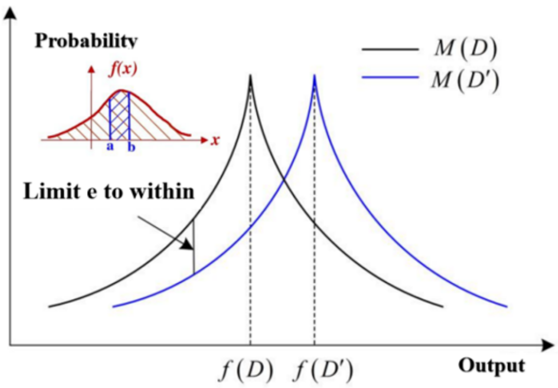

Fig. 2.2: The probability of output based on the randomized algorithm in a set of adjacent data sets

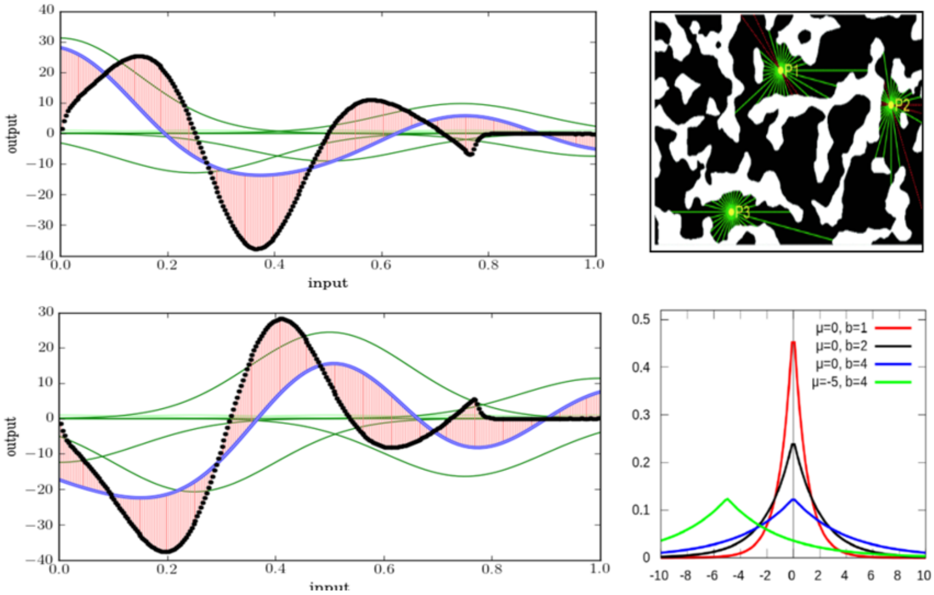

Fig. 2.3: Laplace probability density derived distribution function

2.3. Lift-Apriori-DP new joint algorithm control model. In the Apriori-DP algorithm, the confidence level of the rules is calculated based on the frequency class item set, and the rules that satisfy the confidence level limit boundary value are screened as strong correlation degree rules, but the resulting correlation degree rules are not all correlated, and some of them are invalid correlation degree rules, and the existence of a large number of invalid rules affects the usability of the algorithm when the privacy pre-estimation is reduced. The flow control of the algorithm is shown in Table 2.2.

Figure 2.4 shows the mining results for a data set of 100 data sets extracted for an English score under different privacy pre-estimates, mainly the change of the number of invalid relevance rules in the mining results. As shown in Figure 4, as the privacy pre-estimation decreases, the ratio of invalid relevance rules to the total number of rules in the extraction results gradually increases. In addition, as the privacy pre-estimation decreases,

Table 2.2: Apriori Algorithm Code

Algorithmic Control Flow
Inputs: set of data sets, level of support, level of confidence, privacy pre-estimation
Output: Frequency class item sets and strong correlation degree rules
1. scan the entire set of achievement data sets, k=1, to produce a candidate 1-item set.
2, add noise to the support level of candidate 1-item groups, and filter to produce frequent 1-item groups that satisfy the support level limit boundary value.
3, Self-connection, such that k=k+1, produces candidate k-item groups from frequent (k-1)-item groups.
4, clip, add noise to the support level of candidate k-item groups, and filter to produce frequent k-item groups that satisfy the support limit boundary values.
5, the frequency k-item group is not empty, go to 3.
6, the set of strong association degree rules that satisfy the confidence level is generated from the set of frequency class items.

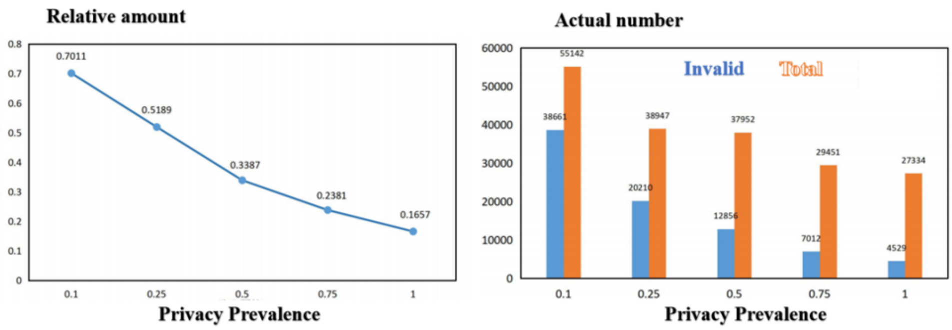

Fig. 2.4: Privacy pre-estimation of correlation metrics in extraction results

the number of correlation rules obtained from mining grows, and the number of invalid correlation rules in them also increases. For the associativity rule mining algorithm based on the protection of differential-degree privacy, the number of invalid associativity rules increases as the noise perturbation increases, and the large number of invalid associativity rules reduces the utility value of the algorithm.

In summary, based on the Apriori-DP algorithm under the framework of the degree of support a degree of confidence, the number of invalid correlation degree rules in the extraction results increases when the privacy degree pre-estimation decreases, and the large number of invalid correlation degree rules reduces the utility value of the algorithm. As a result, boosting degree is introduced to screen and remove the invalid rules that do not meet the relevance. In this chapter, the existing Apriori-DP algorithm is improved by introducing the enhancement degree, and the correlation rules of mining are filtered by the limit boundary value of the support degree, the limit boundary value of the confidence degree, and the limit boundary value of the enhancement degree, and the correlation rules that do not satisfy the relevance are deleted, so that the correlation rules obtained from mining are all effective, and thus the algorithm's efficiency can be improved.

3. An experimental analysis of business English performance under Lift-Apriori-DP.. The experimental data set used in this chapter is the big data of business English in one or two colleges and universities in Chengdu City, Sichuan Province, which is obtained through surveys and studies, to establish a big database under the factors of various variables of business English. The Lift-Apriori-DP algorithm is experimented with the big database as an example, and the experimental results are analyzed. The experimental environment in which the Lift-Apriori-DP algorithm is realized is shown in Table 3.1.

The Lift-Apriori-DP algorithm is validated on different sets of data sets (randomly divided into three groups) to verify the computational effectiveness and accuracy values of the improved algorithm by analyzing the number

Table 3.1: Business English Lift-Apriori-DP Experimental Environment

Type of environment	causality	
host processor	Intel i7	
C-PU	8-core	
random access memory (RAM)	4-GB	
operating system	Windows-10 Flagship	
agglomeration environment	PyCharm-3.1+python-3.7	
Algorithm implementation language	Python	

Table 3.2: Business English Lift-Apriori-DP Experimental Environment

Data set	Frequency k-term group	Level of support	Confidence level (math.)	Privacy pre-estimation	Evaluation indicators	Running time (s)	Number of association degree rules
1	3	0.04	0.7	0.5	sup-conf	0.0249	29
					sup-conf-lift	0.0232	16
2	3	0.04	0.7	0.5	sup-conf	0.6301	301.4
					sup-conf-lift	0.5599	191.4
3	4	0.02	0.7	0.5	sup-conf	273.03	14052.3
					sup-conf-lift	236.85	8832.1

of correlation degree rules extracted and the running time. In the experiments, since the differential degree privacy implementation mechanism involves randomized noise generation when adding noise, the experiments are repeated five times for each set of data and the average value is taken as the experimental result for analysis. The results of the algorithm are analyzed by analyzing the running time of the Apriori-DP algorithm using the degree of support and confidence and the Lift-Apriori-DP algorithm after the introduction of the lifting degree, as well as the number of correlation degree rules generated. The experiments are conducted mainly to compare the extraction results under the same privacy pre-estimation. The privacy degree pre-estimation is set to 0.5 and the minimum lifting degree is 1. The experimental results are shown in Table 3.2.

In Table 3.2, sup-conf denotes the Apriori-DP algorithm with support degree and confidence degree limit boundary values. Sup-conf-lift denotes the Lift-Apriori-DP algorithm with support degree confidence degree and lift limit boundary values, and the experimental results are analyzed as follows. First, the comparison of the number of association degree rules in the extraction results: under the same data set, the Apriori-DP algorithm after the introduction of the lifting degree filters out the invalid association degree rules than the Apriori-DP algorithm consisting of the degree of support and the degree of confidence, so that the output results of the Lift-Apriori-DP algorithm are positively correlated association degree rules, which are valid association degree rules, thus improving the usability of the algorithm. Secondly, for the comparison of the running time of the algorithm: although the Lift-Apriori-DP algorithm increases the computation of lift after the introduction of lift, the overall running time of the Lift-Apriori-DP algorithm is shorter than that of the Apriori-DP algorithm composed of the support degree and the confidence degree. The main reason is that by reducing the number of invalid rules, the length of the output relevance rule list is reduced and the running time is shortened when the set of data increases, the gap between the Apriori-DP algorithm and the Lift-Apriori-DP algorithm in terms of running time and invalid rules filtered out gradually increases.

The data classes in data set 3 are richer and further experimental analysis is performed on data set T10I4TDB00K. Figure 3.1 shows the comparison of the number of association degree rules extracted under the same privacy degree pre-estimation timing and different boosting degree limit boundary values. Setting the privacy degree pre-estimation to 0.5, the orange dashed line shows the number of all association degree rules mined by the Apriori-DP algorithm, and the blue dashed line shows the change in the number of association degree rules mined by the Lift-Apriori-DP algorithm. When the lift limit boundary value increases, the higher

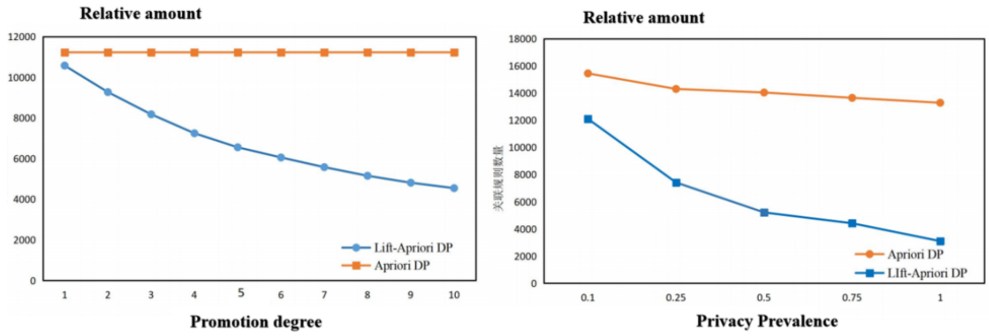

Fig. 3.1: Privacy pre-estimation of correlation metrics in extraction results

chapter1_score	chapter2_score	chapter3_score	chapter4_score	chapter5_score	chapter6_score	chapter7_score	chapter8_score
80	60	40	65	55	50	0	0
0	85	70	65	80	0	0	0
0	90	90	90	90	80	0	0
0	55	80	65	65	80	0	0
60	70	50	90	90	70	0	0
100	90	70	80	80	90	85	83
80	80	90	90	85	70	0	0
100	75	80	80	75	0	64	0
80	75	90	70	70	50	70	0
100	50	70	70	70	60	70	79
80	80	70	90	90	70	73	80
100	55	90	85	90	80	88	92
100	80	70	75	80	80	82	97
0	70	60	85	75	80	88	97
100	70	80	90	70	90	82	78
100	70	100	95	75	80	70	91
80	85	40	55	55	30	73	0
20	40	20	65	55	10	37	2
100	85	80	65	75	80	72	86

Fig. 3.2: Schematic Representation of Business English Raw Performance Data

the relevance of the association degree rules mined by the Lift-Apriori-DP algorithm, the more valid association degree rules obtained gradually decrease. Figure 3.1 shows the variation in the number of invalid correlation degree rules screened out by the Lift-Apriori-DP algorithm with different privacy pre-estimates at a lift of 1. The orange dashed line shows the change in the number of all association degree rules mined by the Apriori-DP algorithm, and the blue dashed line shows the change in the number of invalid association degree rules screened out by the Lift-Apriori-DP algorithm. The number of association degree rules mined by the Apriori-DP algorithm gradually increases when the privacy pre-estimation is decreasing, and the number of invalid association degree rules screened out by the Lift-Apriori-DP algorithm gradually increases after screening by lifting degree.

In addition, to make the data more adequate and convincing in practical use. The actual follow-up test needs to be conducted for the actual situation of business English grades. The follow-up test was conducted and the same instructor taught the Business English course in both academic years, so the student chapter grade data from both academic years were combined. The original set of data sets included a total of 588 students' grade information. The data information mainly includes students' names, genders, student numbers, chapter test scores, etc. Some of the data are displayed as shown in Figure 3.2.

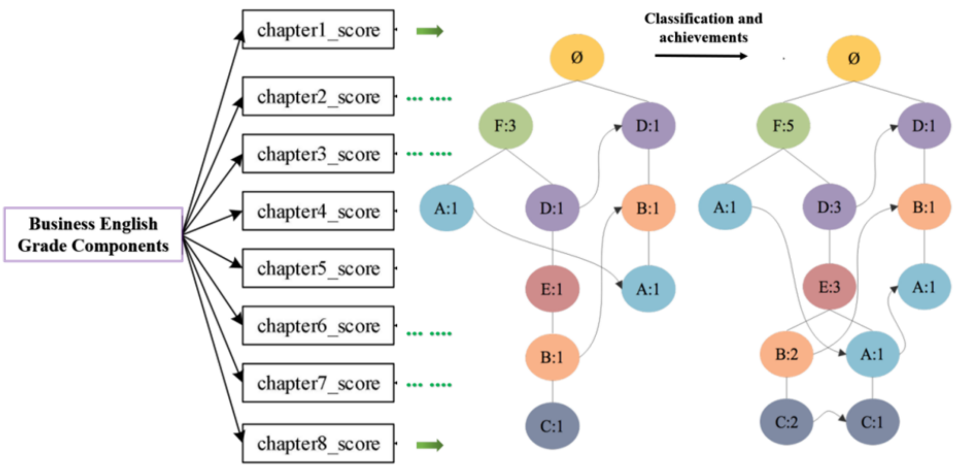

Fig. 3.3: Intelligent Algorithm to Convert Business English Scores Flow

chapter1_score	chapter2_score	chapter3_score	chapter4_score	chapter5_score	chapter6_score	chapter7_score	chapter8_score
A1	D2	E3	E4	E5	E6	E7	E8
A1	A2	A3	A4	A5	A6	C7	A8
A1	B2	B3	A4	B5	A6	C7	B8
A1	A2	B3	A4	E5	A6	C7	A8
A1	B2	A3	C4	B5	C6	C7	A8
B1	E2	E3	E4	E5	E6	E7	E8
A1	A2	C3	B4	B5	A6	B7	B8
A1	E2	C3	C4	C5	D6	C7	C8
B1	B2	C3	A4	A5	C6	C7	B8
A1	E2	A3	B4	A5	B6	B7	A8
A1	B2	C3	C4	B5	B6	B7	A8
A1	C2	B3	A4	C5	A6	B7	C8
A1	C2	A3	A4	C5	B6	C7	A8
E1	E2	E3	D4	E5	E6	E7	E8
A1	B2	B3	D4	C5	B6	C7	B8
A1	B2	D3	C4	B5	A6	C7	B8
A1	B2	E3	A4	B5	D6	C7	B8
A1	C2	A3	B4	B5	D6	C7	B8
A1	C2	C3	B4	C5	D6	B7	A8
A1	C2	B3	D4	D5	C6	D7	D8

Fig. 3.4: Schematic representation of the transformed form of business English performance data

In addition one to eight sections of the performance information is represented by the serial numbers 1 to 8. After the data cleaning, conversion of the preprocessing, with the processing and output process shown in Figure 3.3 data cleaning is the process of processing the original data set set of default values, duplicate values, and anomalous data. In the collected student scores, some students have student information but did not take part in the chapter test, and the chapter test scores are populated with zeros. As a result, there will be certain students with missing course chapter grades in the collected raw data, and to improve the accuracy of the experimental results, the student information with missing certain chapter grades will be deleted. Only the student information with all eight section chapter grades will be retained. After cleaning these invalid data, the remaining 326 students' performance information. After a series of pre-processing of student grades, the data set is described as shown in Figure 3.3.

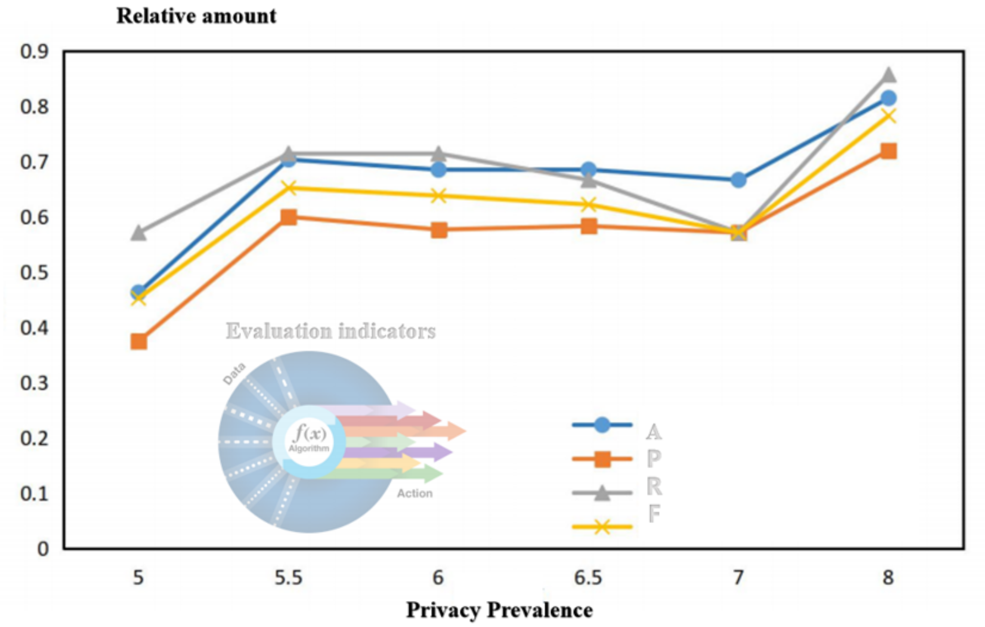

Fig. 3.5: Relative Measures of Correlation Indicators in Actual Measurements

In the actual test results, the data values of each evaluation index according to the correlation degree rule extraction results are plotted as a trend graph, as shown in Figure 3.5. The graph shows that when the privacy preestimation increases, the overall trend of the data accuracy, precision pattern, and recall gradually increases, and the overall value of the comprehensive evaluation index also increases. That is to say, the degree of usability of the data is increasing as the privacy pre-estimation increases. According to the value of the accuracy of the extraction results, when the privacy pre-estimation is chosen to be less than or equal to 5, although the data privacy protection effect is strong, the accuracy of the results will be less than 0.5, and the degree of data usability will decrease. When the privacy pre-estimation is greater than or equal to 7, the accuracy of the data is high, but the protection of the privacy of the data is poor. As a compromise, when the privacy pre-estimation is 5.5, the experimental results are more accurate and usable, i.e., the strong-correlation degree rule extracted has a high degree of accuracy while protecting the privacy of individual students.

4. Conclusions and discussions. This paper investigates the application of data mining based on the correlation degree rule algorithm in business English grades in colleges and universities, aiming at discovering the correlation relationship between students' grades, providing a reference for teaching improvement, and at the same time, protecting students' private information from being leaked. Aiming at the problem that in the Apriori algorithm based on the protection of differential degree privacy, a lot of invalid rules will be mined when the clips are screened according to the degree of support and the degree of confidence, the lifting degree is introduced to improve the Apriori-DP algorithm. The specific conclusions are as follows:

1. Experimental validation is carried out on different sets of data sets, and the experimental results show the effectiveness of the Lift-Apriori-DP algorithm.
2. When the privacy pre-estimation decreases, the correlation degree rules mined by the Apriori-DP algorithm gradually increase, and the number of invalid correlation degree rules screened out by the Lift-Apriori-DP algorithm.
3. In the choice of privacy pre-estimation, too small a privacy pre-estimation may lead to over-protection of data privacy, making the extraction results less accurate. On the contrary, too large a privacy pre-estimation may lead to a weakening of the privacy protection of the data and a decrease in the usability

of the data.

Although the research in this paper has achieved certain results, there are still some limitations and deficiencies such as data sources and data quality, suitable relevance indicators, and so on. In the future, we can further explore and study how to improve the quality and accuracy of data, optimize the algorithm parameter settings, select more appropriate relevance indicators, and enhance privacy protection and data security.

Funding. Education Reform Project of Jiangxi Province-JXJG-23-36-7

REFERENCES

[1] Dengguo, F., Min, Z. & Hao, L. The protection of big data security and privacy. *Journal Of Computing.* **37** pp. 01 (2014)
[2] Zhang, F. An overview of big data technology research. *Communication Technology.* **2014**, 1240-1248 (0)
[3] Han, J., Kamber, M. & Mining, D. Concepts and techniques. *Morgan Kaufmann.* **340** pp. 94104-3205 (2006)
[4] Li, H., Jingling, Z. & Tingting, L. Research on the application of data mining technology in the era of big data. *Modern Information Technology.* **3**, 132-133 (2019)
[5] Romero, C. & Mining, V. a survey from 1995 to 2005. *Expert Systems With Applications.* **33** pp. 01 (2007)
[6] Manek, S., Vijay, S. & Kamthania, D. Educational data mining-a case study. *International Journal Of Information & Decision Sciences.* **8** (2016)
[7] Wang, P., Shi, L., Bai, J. & Others Mining association rules based on apriori algorithm and application. *International Forum On Computer Science-technology &Applications. IEEE.* pp. 141-143 (2010)
[8] Yue, L., Jian, S. & Zhiqi, Q. Research and application of data mining technology based on association degree rule. *Modern Electronic Technology.* **39**, 121-123 (2016)
[9] Solanki, S. & Patel, J. Survey on association rule mining using apriori algorithm[C]. *Fifth International Conference On Advanced Computing & Communication Technologies.* pp. 2327-0659 (2015)
[10] Wang, H. & Liu, P. Application of improved association degree rule algorithm in student achievement alerts. *Computer Engineering And Design 000(00.* **3** pp. 679-682 (2015)
[11] Yuejian, F., Jinzhong, Z., Wen, Z. & Others A review of research on protection algorithms for data mining privacy. *Information Network Security (0.* **2** pp. 6-11 (2017)
[12] Liu-Sheng, H., Miao-Miao, T. & Huang He., A. review of research on cryptographic techniques for protection of privacy in big data. *Journal Of Software.* **26** pp. 04 (2015)
[13] Dalenius, T. Towards a methodology for statistical disclosure control. *Statistik Tidskrift.* **15** pp. 02 (1977)
[14] Privacy, D. Encyclopedia of Cryptography and Security. (0)
[15] Xiong, P., Zhu, T. & XF., W. Differential degree privacy protection and its application[J]. *Journal Of Computing.* **37** pp. 01 (2014)
[16] Li, N. & Wahbeh, H. Qardaji. et al. *PrivBasis: Frequent Itemset Mining With Differential Privacy.* **5**, 1340-1351 (2012)
[17] Wen-yong, G., Ying-jie, W. & Lan, S. et al. Frequent pattern mining with differential privacy based on transaction truncation. *Chin Comput Syst.* **36**, 2583-2587 (2015)
[18] Cheng, X., Su, S., Xu, S. & Others DP-Apriori: A differentially private frequent itemset mining algorithm based on transaction splitting. *Computers & Security.* **50** pp. 74-90 (2015)
[19] Shen, Z. & Yuan, J. A Parallel Differential Degree Privacy Associativity Degree Rule Extractive Algorithm[J]. *Software Guide.* **9**, 65-71 (2017)
[20] Yihui, C., Wei, S., Zhiyong, P. & Others Multi-source data relevance degree rule mining method based on differential degree privacy. *Computer Science.* **45** pp. 36-40 (2018)
[21] Han, Q., Lu, D., Zhang, K. & Others Secure mining of association rules in distributed datasets. *IEEE Access.* **7**, 55325-15533 (2019)
[22] Chen, T. & Long, S. A heuristic algorithm for frequentness class item group mining based on differential degree privacy. *Computer Engineering And Design.* **40** pp. 02 (2019)
[23] Jiang, C., Yang, G., Yunlu, B. & Others Privacy-oriented protection of frequentness class itemset mining type algorithms. *Information Network Security (0.* **4** pp. 73-81 (2019)
[24] Markellou, P., Mousourouli, I., Spiros, S. & Others Using semantic web mining technologies for personalized e-learning experiences. *Proceedings Of The Web-Based Education.* pp. 461-826 (2005)
[25] Shuxun, H. & Zuoyou, Y. Application of relevance degree rule mining theory in teaching degree evaluation. *Software Guide (Educational Technology).* **8** pp. 02 (2009)
[26] Garcia, E., Romero, C., Ventura, S. & Others A collaborative educational association rule mining tool. *Internet & Higher Education.* **14** pp. 02 (2010)
[27] Buldu, A. & Others Data mining application on students' data. *Procedia-Social And Behavioral Sciences.* **2** pp. 02 (2010)
[28] Milicevic, A., Vesin, B., Ivanovic, M. & Others E-learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education.* **56** pp. 03 (2011)
[29] Xiao, Z. The application of relevance degree rule in evaluating distance education teaching degree. *China Distance Education.* **2012** pp. 39-42 (0)
[30] Aher, S. Lobo L M R J. *Combination Of Machine Learning Algorithms For Recommendation Of Courses In E-Learning System Based On Historical Data.* **519** pp. 1-14 (2013)

[31] Hui, G., Qing, Y., Chenggong, J. & Others Research and application of relevance degree rule in performance analysis. *Computer Applications 000(*. **10** pp. 149-151 (2015)
[32] Chen, A. University grade relevance degree rule based on improved Apriori algorithm - An empirical study of a university's letter management program. *Electronic Technology And Software Engineering (0.* **8** pp. 177-178 (2017)
[33] Park, S. Park YB. *Nalysis Of Association Between Students' Mathematics Test Results Using Association Rule Mining.* pp. 1-6 (2018)
[34] Wang, C. Research and application of the Apriori algorithm for association degree rule. *North China Electric Power University (Beijing).* pp. 25-30 (2018)
[35] Kosari, A. & Others Remote sensing satellite's attitude control system: rapid performance sizing for passive scan imaging mode. (Aircraft engineering,2020)
[36] Sharifi, A. Flood Mapping Using Relevance Vector Machine and SAR Data: A Case Study from Aqqala, Iran. (Journal of the Indian Society of Remote Sensing,2020)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: Nov 17, 2023
Accepted: Mar 20, 2024