Risk Prediction Pattern Matching Method of Construction Project Management System in Big Data Era

Qiu-yi Li(✉)

Fuzhou University of International Studies and Trade, Fuzhou 350202, China
lqy13799475957@163.com

Abstract. In view of the influence of the scale of pattern information set on the pattern matching of risk prediction, aiming at improving the performance of pattern matching of risk prediction, this paper puts forward the method of pattern matching of risk prediction of construction project management system in the era of big data. In the era of big data, the central idea of risk prediction pattern matching algorithm is analyzed. Based on the description of risk prediction pattern matching algorithm, the specific implementation steps of risk prediction pattern matching algorithm are designed, the design of risk prediction pattern matching algorithm is completed, and the risk prediction pattern matching process of construction project management system is combined to achieve the risk prediction pattern matching. The experimental results show that, under different experimental platforms, compared with other risk prediction pattern matching methods, the accuracy of risk prediction pattern matching method of construction project management system in the era of big data is higher.

Keywords: Big data era · Construction engineering · Project management system · Risk prediction model

1 Introduction

With the rapid development of computer technology, construction project management system has become an important guarantee of social development. There are many important construction project management information stored, transmitted and processed in the construction project management system, such as macro-control decision-making, commercial economic information, bank funds transfer, stock and securities, energy and resource data, scientific research data and other important information, which will inevitably attract all kinds of human attacks from all over the world (such as information leakage Information stealing, data tampering, data deletion, computer virus, etc. [1, 2]. At the same time, the construction project management system will also be tested by flood, fire, earthquake, electromagnetic radiation and other aspects. Construction project management system has become one of the serious social problems. Construction project management system is a comprehensive discipline involving computer science, network technology, communication technology, cryptography technology, information security technology, applied mathematics, number theory,

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 347, pp. 411–424, 2021.
https://doi.org/10.1007/978-3-030-67871-5_37

information theory and other disciplines [3]. It mainly refers to the protection of the hardware, software and the data in the construction project management system from accidental or malicious reasons, damage, change and leakage, continuous and reliable operation of the system, and uninterrupted network services. Obviously, with the further development of the network, the threat brought by the construction project management system will increase day by day.

At present, there are also individuals and research organizations, mainly represented by Santonatos and Mike Fisk, who actively carry out the research on risk prediction pattern matching methods. China has also stepped up research in this area, such as Wang Yongcheng of Shanghai Jiaotong University, Song Hua of Tsinghua University, and other universities such as Hebei University, northwest Polytechnic University, Harbin Polytechnic University, Nanjing Normal University, etc. It can be seen from the relevant materials and literatures that at present, the research on the risk prediction pattern matching method for the construction project management system still stays in the single pattern matching method, while the research on the multi pattern matching method mainly focuses on the method overview, testing and some corresponding improvements on the existing methods [4]. Although these improved methods have achieved some results, the overall effect is not very ideal, mainly because the method speed is limited by the number of intrusion rules or the space consumption of the implementation method is too large, so the practicability of the method used in the construction project management system is not strong. At the same time, it can also be seen that it is difficult to propose a new risk prediction pattern matching method. Therefore, since adopting the multi pattern matching method for management, the multi pattern matching method introduced by the construction project management product has not changed much, and the main methods still continue to be used, such as the famous snort construction project management system with open source code AC or WM method is adopted; WM method is adopted for easy guard construction project management system; AC-BM method is adopted for Tiantian construction project management products of Beijing Qiming Xingchen company and management series products of Lenovo online.

As the core technology of construction project management system, risk prediction pattern matching technology has always been the focus of the industry. According to the statistics of relevant researchers, in the filtering and detection of the construction project management system, the pattern matching module takes up 70% of the execution time and 80% of the program instructions of the system; about 30% of the risk problems are caused by the low efficiency of data package detection in the construction project management system [5, 6]. With the increase of network bandwidth and the expansion of matching pattern set, the performance of pattern matching method has become the bottleneck of network security system. Therefore, further study and improvement of pattern matching technology is of great significance to improve network fluency, performance of network security system and security of network security system.

In this context, this paper proposes a risk prediction pattern matching method for construction project management system. This paper analyzes the core idea of the risk prediction pattern matching algorithm, designs the risk prediction pattern matching algorithm, and combines the management system to realize the risk prediction pattern matching. In order to verify the effectiveness of the proposed method, an experimental simulation experiment was conducted. Experimental results show that the proposed method is more accurate.

2 Design of Risk Prediction Pattern Matching Method

2.1 Improved Design Risk Prediction Pattern Matching Algorithm

In the construction project management system, the risk prediction pattern matching algorithm is a widely used pattern matching algorithm. One of the most important characteristics of the risk prediction pattern matching algorithm is that in the process of matching the risk prediction pattern strings, many useless risk prediction patterns can be skipped, that is, no useless risk prediction patterns are matched. Through this kind of skip matching, we get a higher execution efficiency [7]. Some experimental data show that the matching speed of risk prediction pattern matching algorithm is about 3–5 times that of other matching algorithms.

The central idea $P(pattern)$ of risk prediction pattern matching algorithm is to assume that the length of risk prediction pattern is m. The left most risk prediction model *pattern* and the left most risk prediction model of the shilling risk prediction model string $T(text)$ are aligned, and then the last risk prediction model $t(m)$ is compared with its corresponding risk prediction model in *text*, that is, matching from right to left. When a mismatch is found, the algorithm adopts two heuristic rules, which are heuristic rules It is: bad risk prediction pattern (BC) rule and good suffix (GS) rule to calculate the moving distance of the pattern risk prediction pattern string, and realize the leaping ergodic matching [8].

Description of bad risk prediction model (BC) Rules.

When a risk prediction mode in the mode risk prediction mode string P is different from a risk prediction mode in the text risk prediction mode string t, a bad risk prediction mode appears. The risk prediction mode matching algorithm moves the mode risk prediction mode string to the right, makes the most right corresponding risk prediction mode in the mode relative to the bad risk prediction mode, and then continues to match. The move function is as follows:

$$BadChar(c) = \begin{cases} m \\ m - j \end{cases} \tag{1}$$

After aligning *pattern* with *text* the left, there are two possibilities for comparing the results of the last risk prediction model in *text* and its corresponding risk prediction model in *pattern*.

The first situation is that $t(m)$ does not appear in any of the *pattern*, so we can move m risk prediction modes to the right of *pattern*, and then compare the last risk prediction mode with the corresponding risk prediction mode ($t(2m)$) in *pattern*.

The second case is: if $t(m)$ is the j risk prediction mode in *pattern*, then we can move *pattern* M-J risk prediction mode to the right.

Good suffix GS rule description.

If a good suffix has been matched, and there is another same suffix in the pattern, GS rule considers the matched situation and determines a new moving distance, which is only related to pattern risk prediction pattern string P.

The steps of risk prediction pattern matching algorithm are as follows:

Step 1: preprocessing, the algorithm moves the pattern risk prediction pattern string to the right as far as possible according to the two pre calculated arrays. The calculation of $Skip[]$ risk prediction model and risk prediction model $Shift[]$, represent BC rule and GS rule respectively.

Step 2: compare the text risk prediction model string and model risk prediction model string one by one from right to left, and continue to compare the single risk prediction model matching. If it reaches the far left of the pattern risk prediction pattern string, the matching occurs successfully and the output is output; if the risk prediction pattern mismatch occurs, the third step is turned.

Step 3: take the array corresponding $Skip[]$ and $Shift[]$ to the mismatch risk prediction mode and the largest value in the array as the moving distance, and move the mode risk prediction mode string to the right. If the end of the text risk prediction mode string has been reached, the algorithm exits; otherwise, it will go back to the second step.

The first mock exam algorithm is the first mock exam algorithm to search for a single pattern risk prediction pattern in building engineering project management system. In a single pattern risk prediction pattern string matching algorithm, the risk prediction pattern matching algorithm is generally considered to be the best performance. There are many keyword patterns to match in content filtering and detection, so the risk prediction pattern matching algorithm needs to match each pattern separately [9]. The time and space complexity of the preprocessing stage of the risk prediction pattern matching algorithm is $O(m+n)$, the time and space complexity of the search stage is $O(mn)$, the worst-case comparison times of the non periodic pattern search is 3 N. The worst time complexity of risk prediction pattern matching algorithm is $O(mn)$, and the best time complexity is $O(n)$. The complexity of directly using risk prediction pattern matching algorithm to match multi-mode risk prediction pattern strings is $O(kn)$. Next, through the risk prediction pattern matching process of the construction project management system, the risk prediction pattern matching is realized.

2.2 Matching Risk Prediction Model

In the construction project management system, the matching of risk prediction pattern is to compare the information collected in the system with the rules in the existing pattern database, so as to judge whether the risk prediction pattern matches. The risk prediction pattern matching process of the construction project management system is shown in Fig. 1.

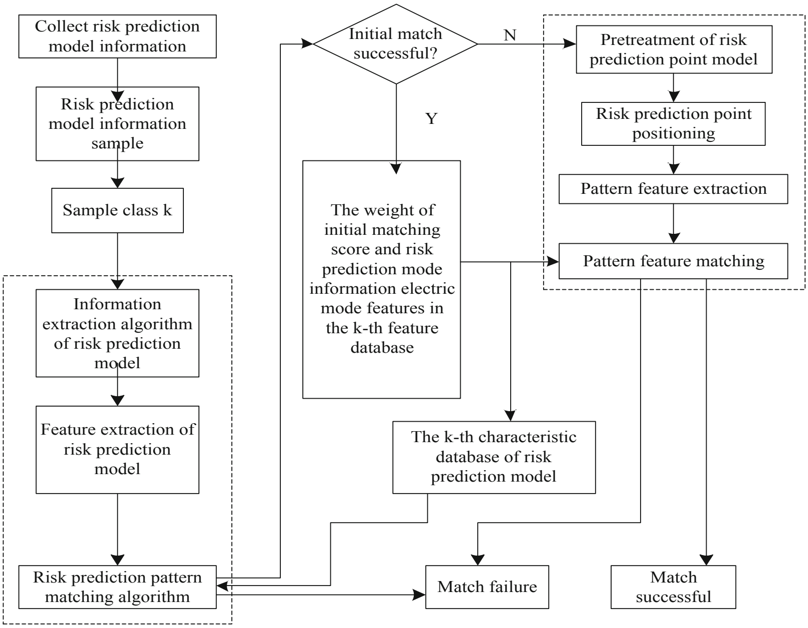

Fig. 1. Risk prediction pattern matching process

At present, the method of risk prediction pattern matching has become one of the most commonly used methods in the field of construction project management, which has been proposed for a long time. However, the survey shows that when using the risk prediction pattern matching algorithm in the era of big data, the pattern matching algorithm determines the detection efficiency and resource occupancy of the construction project management system to a large extent, and the current increase of network speed and network application makes the data in the network increase rapidly, which makes the pattern matching algorithm more and more important. It is precisely because For this reason, the research and exploration of the optimization method of pattern matching algorithm has never stopped, which is becoming more and more popular in today's network security [10, 11].

To sum up, in the era of big data, this paper analyzes the central idea of the risk prediction pattern matching algorithm, designs the specific implementation steps of the risk prediction pattern matching algorithm based on the description of the risk prediction pattern matching algorithm, completes the design of the risk prediction pattern matching algorithm, combines the risk prediction pattern matching process of the construction project management system, and realizes the risk prediction pattern Match of [12, 13].

3 Comparative Analysis of Experiments

3.1 Build Experimental Platform

Experiment is used to verify the performance of risk prediction pattern matching method of construction project management system in the era of big data. On the experimental platform of Table 1, the matching method of risk prediction pattern of construction project management system in big data era is realized on CPU and GPU respectively.

Table 1. Parameters of experimental platform

Platform	Main configuration	Model
Platform 1 Platform 2	CPU	Inteli5-2300
	Memory	2.8 GHz 4G
Platform	GPU	GTX570
	Memory	1 GB
	CUDAVersion	3.2
	OS	Windows7

The experimental model set adopts the URL collected from the construction project management system in 2012, which is divided into five groups, with the scale of 20000, 40000, 80000, 160000 and 320000, as shown in Table 2. 12 MB URL of text data collected in the same way. The hit times in the table refer to the times that the URLs in the URL set can match in the total text data. Because there are duplicate URLs in the text data, the hit times include repeated cumulative URL matches. Taking URL set 1 as an example, there are 20000 URLs in the URL set, and 283256 URLs in URL set 1 are hit in all URL text data. The length of URLs in all 5 URL sets is 8–15 risk prediction modes.

Table 2. Test set

Test set	Number	Hit counts	Text set size
Test set 1	2000	283256	12 MB
Test set 2	4000	515649	12 MB
Test set 3	8000	994134	12 MB
Test set 4	160000	1921158	12 MB
Test set 5	320000	3966726	12 MB

3.2 Analysis of Experimental Results

As the WM based risk prediction pattern matching method of construction project management system has the characteristics of high matching efficiency and less memory, it is one of the classic multi pattern matching methods. In order to evaluate the risk prediction pattern matching method of construction project management system in the era of big data, this paper compares the model matching method of risk prediction of construction project management system based on WM algorithm. Due to the different requirements of platform 1 and platform 2 for the operation environment, this experiment will match the risk prediction pattern matching method of the construction project management system in the era of c-big data and the risk prediction pattern matching method of the construction project management system based on WM algorithm on the experimental platform 1, and match the risk prediction pattern matching method of the construction project management system in the era of g-big data and G- Based on the WM algorithm, the risk prediction pattern matching method of the construction project management system performs the matching task on the experimental platform 2. Four matching methods are used on their own platforms, and five sets of URL sets are used as pattern sets to match with the total data sets.

Figure 2 shows the preprocessing time comparison of four algorithms on five sets of URLs.

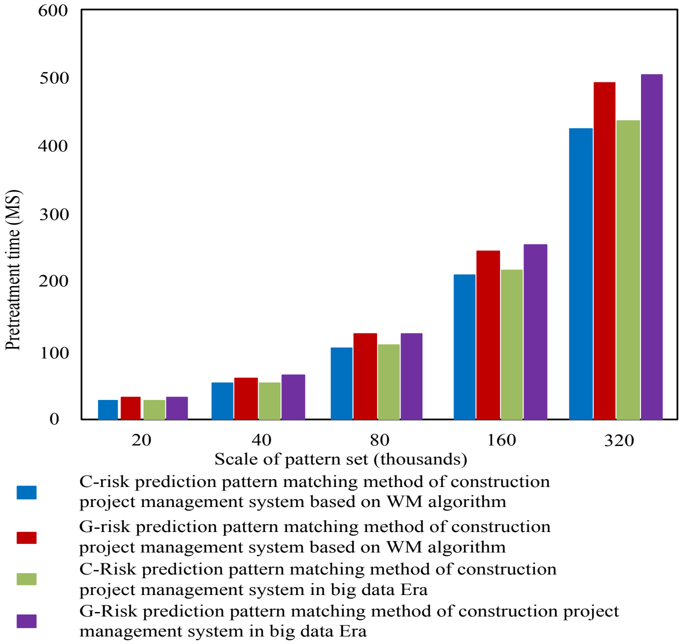

Fig. 2. Comparison results of pretreatment time

From the experimental results in Fig. 2, it can be seen that the preprocessing time of c-risk prediction pattern matching method of construction project management system based on WM algorithm on five sets of URLs is 26.528–423.037 ms, the preprocessing time of c-risk prediction pattern matching method of construction project management system in the era of big data is 27.114–435.947 ms, and the preprocessing time of c-risk prediction pattern matching method of construction project management system in the era of big data is 27.114–435.947 ms The preprocessing time of test pattern matching method is 2.2%–3.05% longer than that of c-wm-based risk prediction pattern matching method of construction project management system; the preprocessing time of g-wm-based risk prediction pattern matching method of construction project management system and g-big data-based risk prediction pattern matching method of construction project management system is longer than that of each other The original algorithm on platform 1. The preprocessing time of G-Risk prediction pattern matching method of construction project management system based on WM algorithm on five sets of URLs is 31.257–495.184 ms, that of G-Risk prediction pattern matching method of construction project management system in the era of big data is 32.116–504.996 ms, and that of G-Risk prediction pattern matching method of construction project management system in the era of big data is 32.116–504.996 ms It is 1.98%–2.75% more than that of G-Risk prediction pattern matching method of construction project management system based on WM algorithm, and the preprocessing time of G-Risk prediction pattern matching method of construction project management system in the era of big data is 15.84%–18.45% more than that of c-risk prediction pattern matching method of construction project management system in the era of big data.

The preprocessing time of the matching method on platform 2 is longer than that of the original algorithm on platform 1. Because there are fewer logic operation components on platform 2, the preprocessing of the algorithm in the early stage is carried out on platform 1, and the memory of platform 1 and platform 2 cannot be addressed uniformly, so the preprocessed data structure on platform 1 can only be transferred to the memory of platform 2 one by one. Compared with the original algorithm based on platform 1, the algorithm based on platform 2 needs to transfer data time.

The matching time comparison results of different matching methods are shown in Fig. 3.

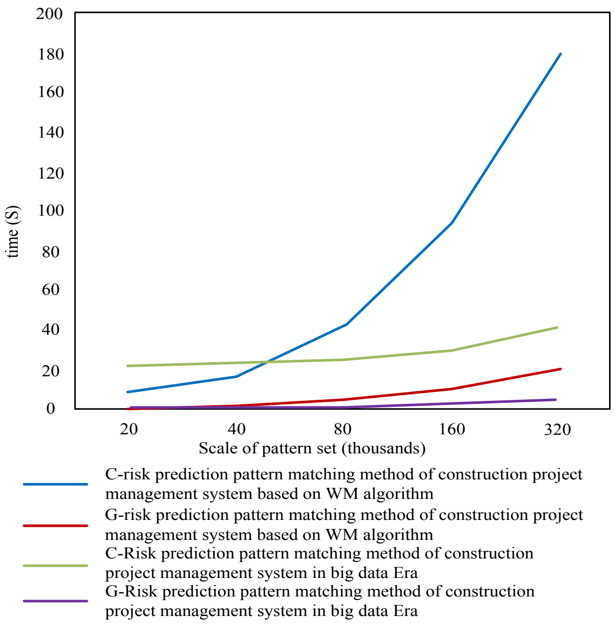

Fig. 3. Matching time comparison results

From the experimental results in Fig. 3, it can be seen that (a) on the large pattern set (more than 80000 items), the performance of the risk prediction pattern matching method of the construction project management system in the era of c-big data is better than that of the risk prediction pattern matching method of the construction project management system based on the WM algorithm, and the larger the pattern set, the better the performance. This can be attributed to two reasons: one is that the conflict rate of c-wm-based risk prediction pattern matching method of construction project management system increases with the increase of the scale of pattern set, resulting in the reduction of algorithm efficiency; the other is that the larger the pattern set, the more obvious the advantage of c-big data Era construction project management system risk prediction pattern matching method in the structure parallelism.

(b) Among the two matching methods based on platform 2, the performance of G-Risk prediction pattern matching method of construction project management system and G-Risk prediction pattern matching method of construction project management system based on WM algorithm are better than their respective original algorithms, and the efficiency of risk prediction pattern matching method of construction project management system in g-big data Era is far away Compared with the g-wm based risk prediction pattern matching method, the larger the pattern set, the more significant the advantage. In addition to the two reasons mentioned in (a), there are three reasons for this result: G-Risk prediction pattern matching method of construction project management system in the era of big data makes full use of the multi-layer storage structure of platform 2 to improve data processing speed, while due to the structure problem of matching method, G-Risk prediction pattern matching method of construction project management system based on WM algorithm It is difficult to use platform 2 high-speed on-chip memory shared memory to speed up processing speed on large mode set.

The comparison results of memory consumption of different matching methods are shown in Fig. 4.

From the experimental results in Fig. 4, it can be seen that with the increase of the number of URLs, the memory occupied by different matching methods increases correspondingly. Moreover, the memory occupied by the matching method of risk prediction mode of construction project management system in the era of c-big data is less than that of the matching method of risk prediction mode of construction project management system based on WM algorithm, and the memory occupied by the matching method of risk prediction mode of construction project management system in the era of g-big data is less than that of the matching method of risk prediction mode of construction project management system based on WM algorithm Match method. This is because in addition to suffix table and prefix table, c-risk prediction pattern matching method of construction project management system based on WM algorithm also needs to build jump table, while in the era of c-big data, risk prediction pattern matching method of construction project management system only needs to build multiple extended bloomfilters based on pattern set, and does not need extra memory space.

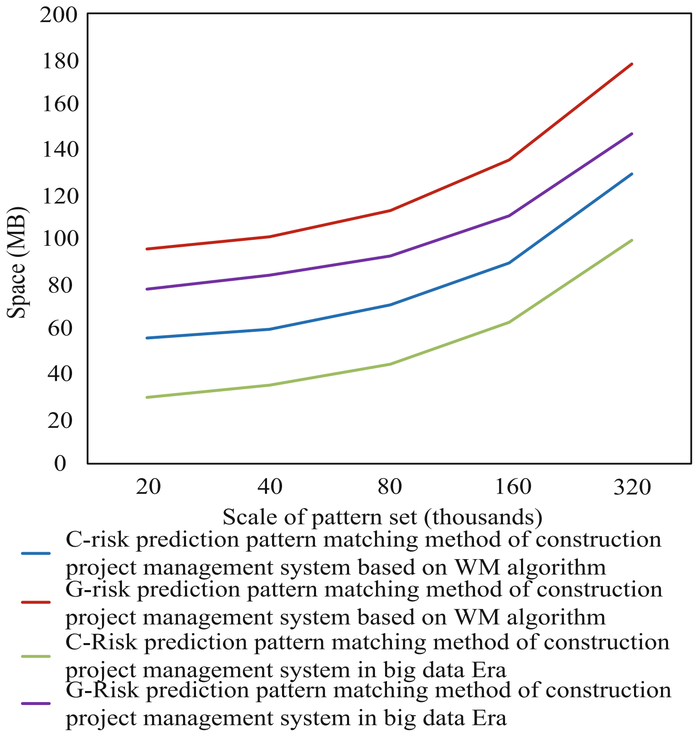

Fig. 4. Comparison results of occupied memory

The throughput comparison results of different matching methods are shown in Fig. 5.

From the experimental results in Fig. 5, we can see that with the increase of the size of the URL pattern set, the throughput of different matching methods decreases gradually, and the efficiency of all matching methods tends to bottleneck. This is because the pattern set is close to the upper limit of GPU memory. In contrast, in the era of g-big data, the throughput of risk prediction pattern matching method of construction project management system is always higher than that of other matching methods, and it is 2–5 times higher than that of risk prediction pattern matching method of construction project management system based on WM algorithm. In the era of c-big data, when the scale of pattern set is less than 40000, the throughput of risk prediction pattern matching method of construction project management system based on WM algorithm is lower than that based on c-big data. When the scale of pattern set increases to 80000 or more, the throughput of risk prediction pattern matching method of construction project management system in the era of c-big data is higher than that of risk

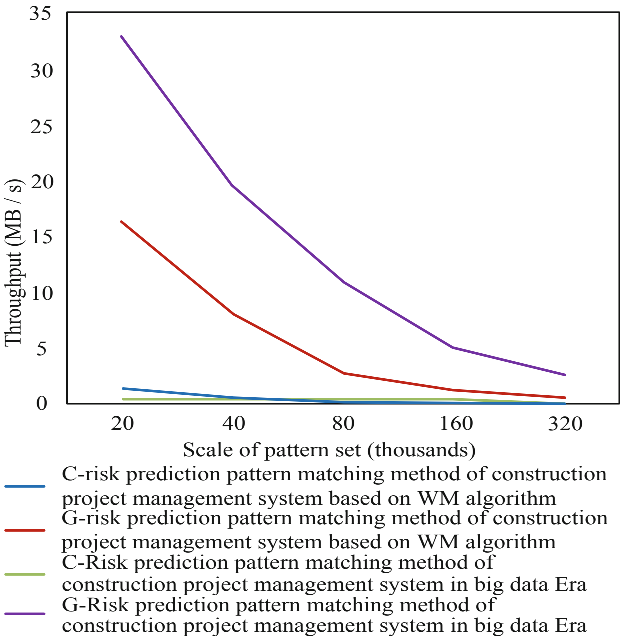

Fig. 5. Throughput comparison results

prediction pattern matching method of construction project management system based on WM algorithm. This is because of the small scale of pattern set, the conflict rate of c-risk prediction pattern matching method of construction project management system based on WM algorithm is smaller, and the efficiency is higher than that of c-risk prediction pattern matching method of construction project management system in the era of big data. With the rapid increase of pattern set, c-risk prediction pattern matching method of construction project management system based on WM algorithm is rushed The burst rate increases and the matching efficiency decreases.

This paper also compares the acceleration ratio (serial time/parallel time) of the risk prediction pattern matching method of the construction project management system in the era of g-big data on different pattern sets with the risk prediction pattern matching method of the construction project management system based on the WM algorithm, as shown in Fig. 6.

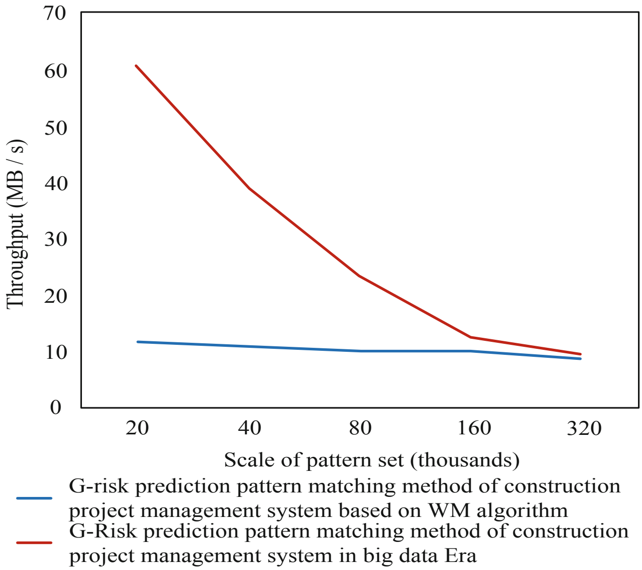

Fig. 6. Acceleration ratio comparison results

From the experimental results in Fig. 6, it can be seen that the acceleration ratio of the risk prediction pattern matching method of the construction project management system in the era of g-big data is always higher than that of the risk prediction pattern matching method of the construction project management system based on the WM algorithm. At the same time, it decreases with the increase of the pattern set, and gradually matches with the risk prediction pattern matching formula of the construction project management system based on the WM algorithm The acceleration ratio of the method tends to be the same. This is because the GPU's memory limit is close at this time. When the memory is large enough, the difference between the acceleration ratio of the two is very obvious. In the era of g-big data, the acceleration ratio of risk prediction pattern matching method of construction project management system can reach 60 times in the best case and nearly 10 times in the worst case. The acceleration of g-wm-based risk prediction pattern matching method for construction project management system is only about 12 times of the best case. This fully shows that in the era of c-big data, the degree of structural parallelism of risk prediction pattern matching method of construction project management system is far greater than that of c-wm based risk prediction pattern matching method of construction project management system, which is more suitable for parallel pattern matching [14, 15].

Based on the above experimental results, the performance of risk prediction pattern matching method in construction project management system is far better than other matching methods, no matter on platform 1 or platform 2.

4 Conclusion

In this paper, the risk prediction pattern matching method of construction project management system in the era of big data is proposed. In the background of big data era, the risk prediction pattern matching algorithm is designed. Combined with the risk prediction pattern matching process, the risk prediction pattern matching of construction project management system is realized. The results show that the risk prediction pattern matching method of construction project management system in the era of big data has higher accuracy.

Due to the urgency and importance of big data processing, big data technology has been highly valued in the global academic circle. In this paper, the pattern matching of risk prediction of construction project management system has been completed under the background of big data, but the implementation of project management after matching still needs further research, which will also be the focus of future research.

Acknowledgements. This research is funded by (Research on the application of cloud computing technology in the construction project management system in the era of big data, with No. JAT190908).

References

1. Hassan, G.M.: Discontinuous and pattern matching algorithm to measure deformation having discontinuities. Eng. Appl. Artif. Intell. **81**(2), 223–233 (2019)
2. Santiago, Cláudio P., Lavor, C., Monteiro, S.A., Kroner-Martins, A.: A new algorithm for the small-field astrometric point-pattern matching problem. J. Glob. Optim. **72**(1), 55–70 (2018). https://doi.org/10.1007/s10898-018-0653-y
3. Shin, H.S., Turchi, D., He, S., et al.: Behavior monitoring using learning techniques and regular-expressions-based pattern matching. IEEE Trans. Intell. Transp. Syst. **20**(4), 1289–1302 (2019)
4. Golan, S., Kopelowitz, T., Porat, E.: Streaming pattern matching with d wildcards. Algorithmica **81**(5), 1988–2015 (2019). https://doi.org/10.1007/s00453-018-0521-7
5. Korman, S., Reichman, D., Tsur, G., et al.: Fast-match: fast affine template matching. Int. J. Comput. Vision **121**(1), 111–125 (2017). https://doi.org/10.1007/s11263-016-0926-1
6. Wang, L., Zhiwen, Y.U., Guo, B., et al.: Mobile crowd sensing task optimal allocation: a mobility pattern matching perspective. Front. comput. Sci. China **12**(2), 231–244 (2018). https://doi.org/10.1007/s11704-017-7024-6
7. Gandotra, E., Singla, S., Bansal, D., et al.: Clustering morphed malware using opcode sequence pattern matching. Recent Pat. Eng. **12**(1), 30–36 (2018)
8. Ai, L., Ramaswamy, L., Luo, S.: Impact vertices-aware diffusion walk algorithm for efficient subgraph pattern matching in massive graphs. IEEE Access **7**, 44555–44561 (2019)
9. Lu, M., Liu, S., Kumarsangaiah, A., et al.: Nucleosome positioning with fractal entropy increment of diversity in telemedicine. IEEE Access **6**, 33451–33459 (2018)
10. Kociumaka, T., Pissis, S.P., Radoszewski, J.: Pattern matching and consensus problems on weighted sequences and profiles. Theory Comput. Syst. **63**(3), 506–542 (2019). https://doi.org/10.1007/s00224-018-9881-2
11. Asadi, P., Rezaeian Zeidi, J., Mojibi, T., et al.: Project risk evaluation by using a new fuzzy model based on Elena guideline. J. Civ. Eng. Manag. **24**(5), 284–300 (2018)

12. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
13. Shuai, L., Weiling, B., Nianyin, Z., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)
14. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mobile Netw. Appl. **24**(1), 1–4 (2019)
15. Liu, S., Sun, G., Fu, W. (eds.): eLEOT 2020. LNICST, vol. 340. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63955-6