



Article

Method for Mid-Long-Term Prediction of Landslides Movements Based on Optimized Apriori Algorithm

Wenhao Guo ^{1,2} , Xiaoqing Zuo ², Jianwei Yu ^{1,3,*} and Baoding Zhou ^{1,4} 

¹ Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation & Shenzhen Key Laboratory of Spatial Smart Sensing and Services & Research Institute for Smart Cities & Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China

² Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China

³ College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

⁴ Institute of Urban Smart Transportation & Safty Maintenance, Shenzhen University, Shenzhen 518060, China

* Correspondence: jwyu@szu.edu.cn

Received: 31 July 2019; Accepted: 6 September 2019; Published: 11 September 2019



Abstract: In the study of the mid-long-term early warning of landslide, the computational efficiency of the prediction model is critical to the timeliness of landslide prevention and control. Accordingly, enhancing the computational efficiency of the prediction model is of practical implication to the mid-long-term prevention and control of landslides. When the Apriori algorithm is adopted to analyze landslide data based on the MapReduce framework, numerous frequent item-sets will be generated, adversely affecting the computational efficiency. To enhance the computational efficiency of the prediction model, the IAprioriMR algorithm is proposed in this paper to enhance the efficiency of the Apriori algorithm based on the MapReduce framework by simplifying operations of the frequent item-sets. The computational efficiencies of the IAprioriMR algorithm and the original AprioriMR algorithm were compared and analyzed in the case of different data quantities and nodes, and then the efficiency of IAprioriMR algorithm was verified to be enhanced to some extent in processing large-scale data. To verify the feasibility of the proposed algorithm, the algorithm was employed in the mid-long-term early warning study of landslides in the Three Parallel Rivers. Under the same conditions, IAprioriMR algorithm of the same rule exhibited higher confidence than FP-Growth algorithm, which implied that IAprioriMR can achieve more accurate landslide prediction. This method is capable of technically supporting the prevention and control of landslides.

Keywords: prediction of landslides; MapReduce; Apriori algorithm; Three Parallel Rivers

1. Introduction

Geological disasters pose a serious threat to the safety of human life and private property, and landslides are one of the commonest geological disasters. The occurrence of landslides not only seriously threatens the safety of human life and private property, but also severely damages the environment and ecology [1]. One of the feasible ways to comprehensively control landslide hazards is to predict landslide hazards [2].

In accordance with system theory and nonlinear science, many novel techniques have been employed to study the stability of landslides, a range of comprehensive prediction models have been summarized, and landslide prediction criteria have been proposed, including decision tree [3,4], generalized FLAIR model (GFM) [5], artificial neural network [6], object-oriented methods [7] and support vector machine (SVM) [8]. The landslides of different causes were systematically summarized,

and these landslides of different types were comprehensively analyzed to build prediction models for different landslides [9]. However, these prediction models are not applicable to the mid-long-term application of landslide hazards [10]. Furthermore, landslide occurrence is difficult to predict accurately, and the mid-long-term early warning of landslides is hard to achieve [11].

Recently, the big data technology has been increasingly employed for landslide disaster early warning [12,13]. Through the analysis of historical landslide data, the use of big data technology becomes a hotspot for prompt and feasible landslide mid-long-term early warning. The distributed landslide prediction model does not enhance the efficiency of the algorithm. In the study on mid-long-term landslide warning, the operational efficiency of the prediction model is vital to the timeliness of landslide prevention and control [14].

Association rule mining can be used to determine the levels of association among various causes [15]. Association rule mining has been successfully applied to uncover such cause-and-effect relationships in a variety of fields, including causes of occupational accidents [16] and prediction landslide hazards [17]. When the Apriori algorithm runs based on the MapReduce framework, the Apriori algorithm will generate numerous frequent item-sets, significantly affecting the efficiency [18]. To solve this problem, the Apriori algorithm runs based on the MapReduce framework [13,18], whereas it does not optimize the algorithm for landslide data correlation analysis based on the MapReduce framework. In the processing of large-scale landslide data, the algorithm has low efficiency since it should generate many frequent items, which adversely affects the timely prevention and control of landslides [18,19].

To enhance the computational efficiency of the prediction model, the Apriori algorithm was optimized based on the MapReduce framework, and the IAprioriMR algorithm is proposed in this paper. The Three Parallel Rivers area with frequent geological disasters was taken as the study area. The relevant data of landslides in the study area were collected. The feasibility of the algorithm was verified using the IAprioriMR for the mid-long-term early warning of landslide.

2. Method

2.1. IAprioriMR

Apriori is one of the commonest association rule algorithms, and several methods have been derived using the Apriori algorithm [20]. The Apriori algorithm refers to a loop method that uses hierarchical sequential search, also termed as an iterative method of layer-by-layer search to generate frequent item-sets. However, a single host cannot handle a large number of operations. Processing considerable data is now becoming a new trend. Recently, many algorithms have been proposed, which apply the Apriori in the MapReduce computing framework [21]. This pattern is based on MapReduce, which is hereinafter referred to as AprioriMR. AprioriMR produces all candidate items one time in the Map task, and the form of each candidate itemset is a product with $\langle k, v \rangle$ form. Subsequently, according to the differences of Key, the candidate items are classified and counted by Reduce with the minimum support to screen the high frequency candidate item-sets [22].

The optimized conventional Apriori algorithm can fit the MapReduce framework for distributed parallel computing. However, because it produces all candidate item-sets one time in the Map task, there will be a large memory footprint, significantly decreasing the operating efficiency [23]. Thus, the counting in the Map stage will be time-consuming, and valuable association rules are difficult to find. The Apriori algorithm imposes a heavy burden to MapReduce [24].

In general, the association rules are expressed as $X \Rightarrow Y$. The definition is as follows: in the presence of X, Y , the large-scale item-sets are defined. $I = \{i_1, i_2, \dots, i_n\}$ is a set of items, and a pattern P is defined as $\{P = \{i_j, \dots, i_k\} \subseteq I, j \geq 1, k \leq n\}$. Given a pattern P , its length or size is expressed as $|P|$, i.e., the number of the singletons it covers. Thus, for $P = \{i_1, i_2, \dots, i_k\} \subseteq I$, its size is defined as $|P| = k$. Besides, given a set of all transactions $T = \{t_1, t_2, \dots, t_m\}$ in a dataset, the support of a pattern P is defined as the number of transactions that P satisfies, i.e., $\text{support}(P) = |\{t_l \in T: P \subseteq t_l\}|$. A pattern P is

considered frequent if $\text{support}(P) \geq \text{threshold}$ [22], in this paper, the threshold of minimum support is 1%. Thus, for each pair $(P_l, \{\text{supp}(P)_1, \text{supp}(P)_2, \dots, \text{supp}(P)_m\})$, the result is a $\{P, \text{supp}(P)\}$ pair, so $\text{supp}(P) = \sum_{l=1}^m \text{supp}(P)_l$. This method, when dealing with big data, due to the large scale of the transaction, will yield a huge candidate set.

This study aimed to enhance the computational efficiency of a novel algorithm (hereinafter referred to as IAprioriMR). It does not yield the whole set of candidate item-sets C for each transaction $t_l \in T$, whereas the subset $c \subseteq C$ consists of patterns of size $|P| = s$. Hence, a set of iterations are required, one per different item-set-size (see Algorithm 1). Subsequently, the dataset is split into different chunks of data, one per mapper, and each mapper is responsible for analyzing each single transaction $t_l \in T$ to generate $P, \text{support}(P)_l$ pairs. Lastly, this algorithm also covers multiple reducers to scale down the computational cost. The major difference between AprioriMR and IAprioriMR lies in the mapper, since IAprioriMR obtains any pattern P of size $|P| = s$ for each sub-database (see Figure 1). Thus, the number of $P, \text{support}(P)_l$ pairs generated by each mapper is lower than that by the AprioriMR algorithm [22].

Algorithm 1. IApriori MR Algorithm

```

begin procedure IAprioriMapper( $t_l, s$ )
  1:  $C = \{\forall P : P = \{i_1, \dots, i_n\} \wedge P \subseteq t_l \wedge |P| = s\}$ 
    // candidate item-sets of size  $s$  in  $t_l$ 
  2:  $\forall P \in C$ , then  $\text{supp}(P)_l = 1$ 
  3: for all  $P \in C$  do
  4:   emit  $\{P, \text{supp}(P)_l\}$  // emit the  $\{k, v\}$  pair
  5: end for
end procedure
// In a grouping procedure values  $\text{supp}_m$  are grouped for
// each pattern  $P$ , producing pairs
 $\{P, \text{supp}_1, \text{supp}_2, \dots, \text{supp}_m\}$ 
begin procedure IAprioriReducer
( $\{P, \{\text{supp}(P)_1, \dots, \text{supp}(P)_m\}\}$ )
  1:  $\text{support} = 0$ 
  2: for all  $\text{supp} \in \{\text{supp}(P)_1, \text{supp}(P)_2, \dots, \text{supp}(P)_m\}$  do
  3:    $\text{support} += \text{supp}$ 
  4: end for
  6:   emit  $\{P, \text{support}\}$ 
end procedure

```

2.2. Algorithm Performance Analysis

The frequent pattern (FP)-growth algorithm, designed to solve the problem of high number of transactions and comparisons, is well-known in the association rule. FP-growth stores the frequent item-sets into a tree structure, requiring data to be scanned just once. Nevertheless, it still faces considerable candidate item-sets since either the larger number of I/O or the high memory requires to store all sets. Based on the FP-Tree structure, many authors have extended the FP-Growth method [20].

Three algorithms were employed for test in the same hardware environment (DELL R730 dual 8 cores), and the Webdoc dataset was adopted as the experimental data, Webdoc dataset is considered by the pattern mining community as the biggest one. Besides, AprioriMR and IAprioriMR operating environments are three virtual nodes on the physical machine. In different instances, following the same association rules, the performances of three algorithms differ significantly (see Figure 2). In the same instances, the number of nodes in MapReduce significantly affects the performance of the algorithm (see Figure 3).

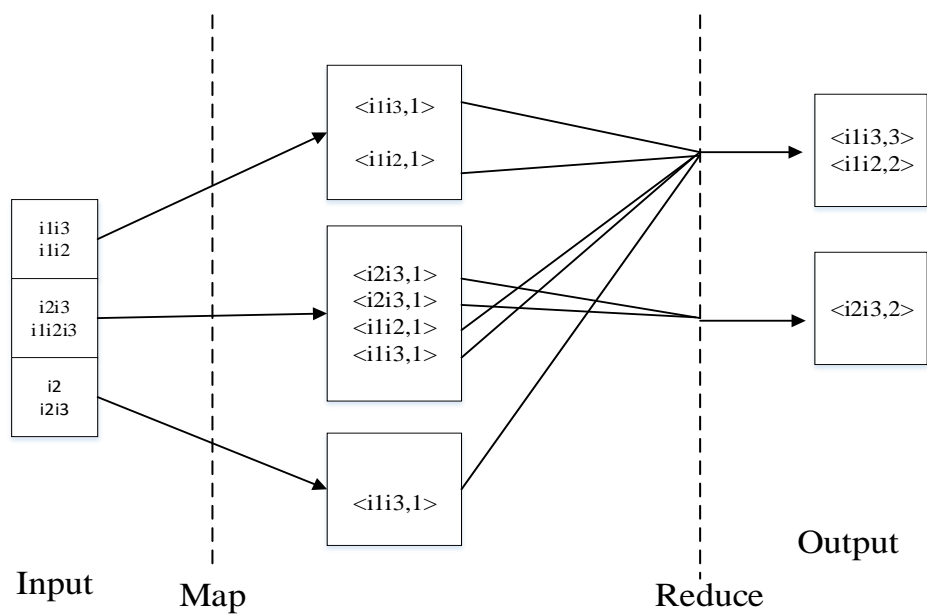


Figure 1. Diagram of the IAprioriMR algorithm.

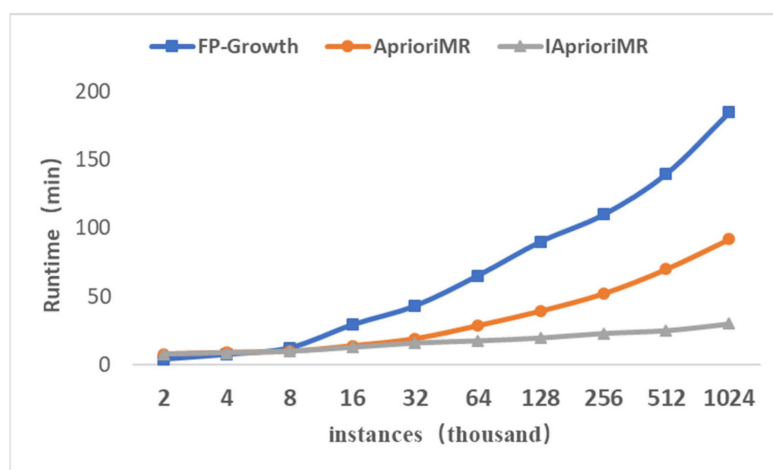


Figure 2. Runtime of different algorithms on the dataset and a number of instances that varies from 2000 to 1,024,000.

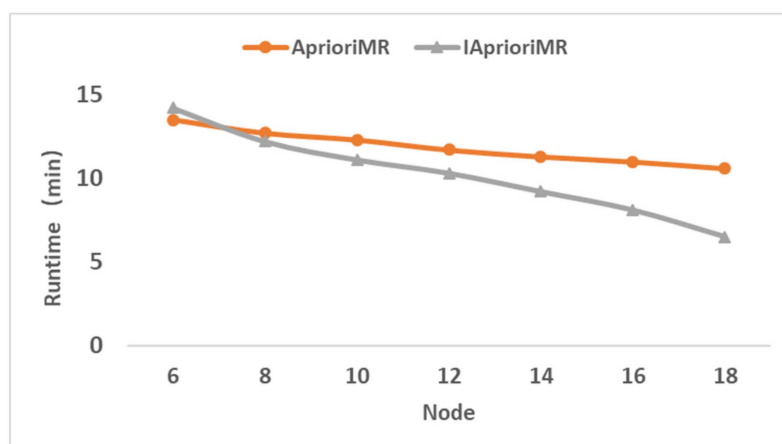


Figure 3. Runtime of nodes from 8 to 20 for different algorithms in 128,000 instances.

According to the results of comparison experiments, in the instance greater than 32,000, the modified IAprioriMR algorithm mode will be significantly optimized as compared with the conventional parallel AprioriMR and FP-growth. When the amount of data is above 1024 thousand, the efficiency of the IAprioriMR algorithm increases by more than 50% compared with that of the AprioriMR algorithm. With the rise in the number of instances, the computational efficiency of the IAprioriMR algorithm increases more significantly. In the meantime, with the increases in the number of nodes, IAprioriMR significantly shortens the processing time of MR and enhances the efficiency in processing large-scale data.

3. Study Area and Data

3.1. Study Area

The Three Parallel Rivers cover the Nujiang River, the Lantsang River, the Jinsha River, and the mountains in the basin at $26^{\circ}03' \sim 29^{\circ}16' \text{ N}$ and $98^{\circ}7' \sim 101^{\circ}19' \text{ E}$ (see Figure 4). The north and northwest have temperate and cold temperate monsoon climate, and the south has a plateau monsoon climate. From the Meili Snow Mountain at 6740 m above sea level to the Nu River at an altitude of nearly 700 m (see Figure 5), the terrain height varies significantly, and the geological structure is complex. Thus, this area is a geological disaster-prone area [25].

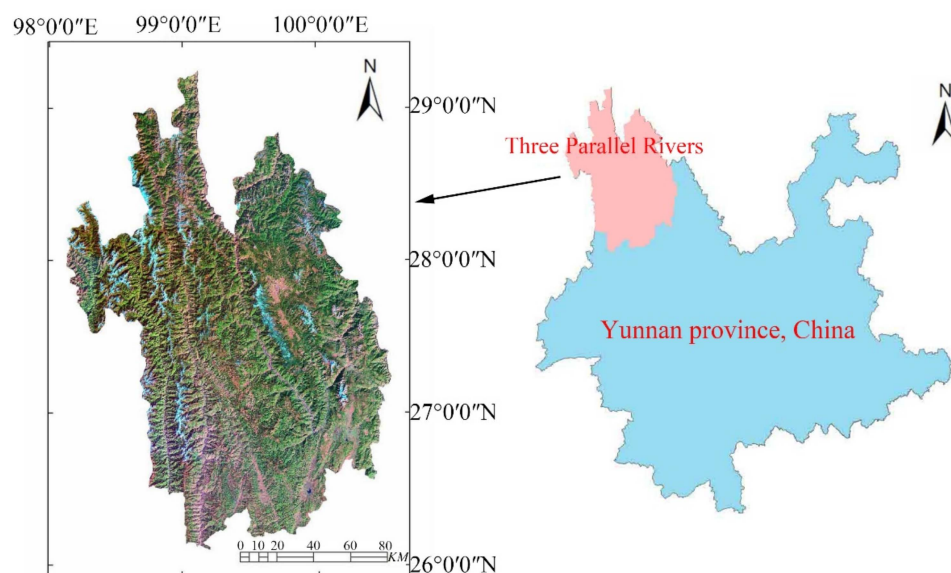


Figure 4. Three Parallel Rivers area location.

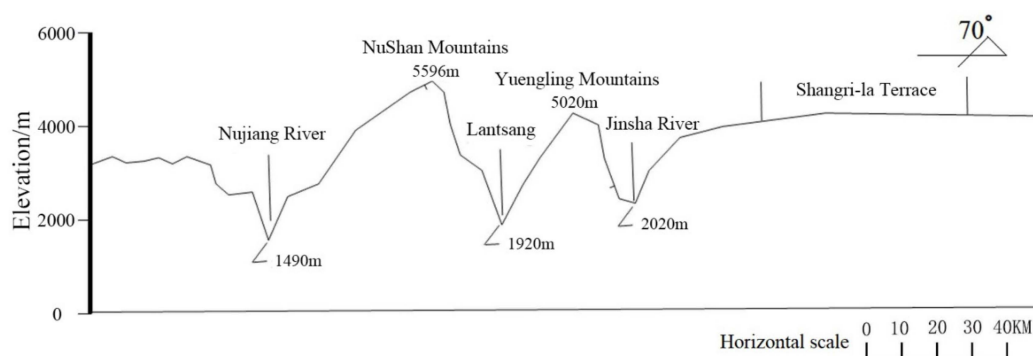


Figure 5. Topographical profile of the Three Parallel Rivers.

The occurrence of landslides has numerous factors. The landslide monitoring data generally refers to the quantitative value of the induced factors (e.g., rainfall and water level) affecting the slope

variation. During the landslide formation and occurrence, there are also many uncertain factors (e.g., human engineering and sudden earthquakes) [26], as well as numerous landslide monitoring data, including graphical data (GIS data), image data (remote sensing data), and relevant geophysical data. The spatial data (e.g., spatial location and spatial relationship) in landslide monitoring data takes up a large proportion, so landslide monitoring data have close correlation.

The landslide monitoring data exhibit the following characteristics: the attributes of the landslide monitoring data exhibit strong and continuous spatial correlation, and the use of the regression algorithm and the fitting algorithm helps build the pattern for analysis; the effect of rainfall on the displacement of landslide is certainly accumulated and delayed, and the occurrence of landslide may be shortly affected by heavy rainfall or attributed to the rainfall accumulated over time.

3.2. Data

The landslide monitoring data from 2000 to 2013 in the Three Parallel Rivers were taken as the experimental data. The total data is 19.6 G, covering a total of 175,000 instances. There are three landslide displacement monitoring stations (the ShangQiaotou, XiangDuo and JuDian landslide monitoring stations), representing the upstream, midstream and downstream areas of the Three Parallel Rivers, respectively.

The Three Parallel Rivers have high rainfall and a clear climate. The rainfall peak in the Three Parallel Rivers area appeared in 2004, and the rainfall has gradually decreased in recent years, as suggested from the rainfall data from 2000 to 2013 (see Figure 6). The groundwater levels in the upstream, midstream and downstream areas of the Three Parallel Rivers area were recorded by ShangQiaotou, XiangDuo and JuDian landslide displacement monitoring stations, respectively. The monitoring station measures the data once a month. Figure 7 shows the annual average rate of change of groundwater level in the area of the Three Parallel Rivers.

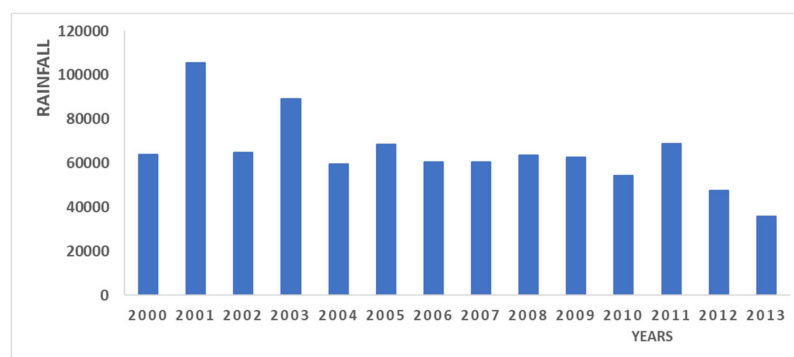


Figure 6. Annual average rainfall in the Three Parallel Rivers from 2000 to 2013.

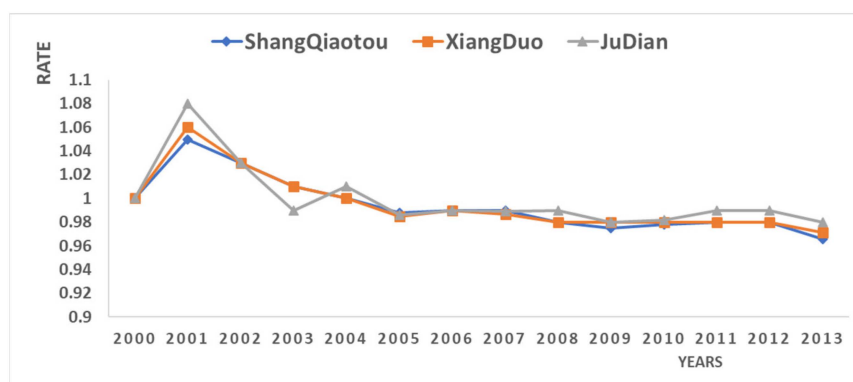


Figure 7. Annual average rate of change of groundwater level in the area of the Three Parallel Rivers.

Because of the special geographical location, geological disasters frequently occur in the Three Parallel Rivers area. Table 1 lists the types of landslide for each landslide hazard, the nature of the landslide, and the maximum rainfall, etc.

Table 1. Information on some landslide disasters in the area of the Three Parallel Rivers.

Name	Longitude	Latitude	Occurrence Time	Elevation Top (m)	Elevation Foot (m)	Landslide Type	Substrate	Stratigraphic Age	Daily Maximum Rainfall (mm)	Maximum Rainfall (mm)
XTBG Landslide	104.449722	26.230278	2005/7/1	1688	1614	retrogressive landslide	soil slopes	P2	120.4	78.3
ZJS Landslide	104.446667	26.196667	2004/5/1	2032	1932	retrogressive landslide	soil slopes	P1	120.4	78.3
DJB Landslide	104.455278	26.226944	2007/6/1	2124	1980	retrogressive landslide	soil slopes	P2	120.4	78.3
XTB Landslide	104.448056	26.230556	2007/8/1	1937	1718	retrogressive landslide	Rock Slope	P2	120.4	78.3
SLG Landslide	104.373833	26.153056	2009/7/1	2215	2134	thrust load caused landslide	soil slopes	P2	120.4	78.3
DCZ Landslide	104.388778	26.167611	2009/7/1	2109	2038	retrogressive landslide	soil slopes	P2	120.4	78.3
ZBZ Landslide	104.380028	26.149194	2008/7/1	2144	2114	composite landslide	soil slopes	P2	120.4	78.3
XBZ Landslide	98.575417	25.184028	2001/4/1	1755	1715	retrogressive landslide	soil slopes	Q4	94.7	47.5

4. Experiment and Analysis

Studies reported that the occurrence of landslide hazards has been associated with the geological environment (e.g., rainfall, geological structure and groundwater distribution) [27]. In the experiment, setting the groundwater level of the monitoring point, the rainfall, the water level of the parallel river in the Three Parallel Rivers, and the cumulative displacement of the landslide monitoring point are the antecedents of the landslide occurrence, and the landslide occurs, which is detectable. The landslide monitoring data from 2000 to 2011 in the Three Parallel Rivers area were taken as the training pattern, and the landslide monitoring data from 2012 to 2013 as the test data [28].

For the convenience of calculation, each group of data was standardized and split into averages and low-value areas [29], covering Under_min (monitoring point low water table), Under_max (monitoring point high water table), Rain_min (small rainfall), Rain_max (large rainfall), River_min (low water level), River_max (high water level), LS1_min (small displacement of the upper bridge monitoring station), LS_max (large displacement of the upper bridge monitoring station), LS2_min (small multi-monitoring station displacement), LS2_max (large multi-monitoring station displacement), LS3_min (small displacement of Jiaodian monitoring station), LS3_max (large displacement of Jiaodian monitoring station), LandslideD_1 (Formation lithology is Gravel soil), LandslideD_2 (Formation lithology is Slate), LandslideD_3 (Formation lithology is Conglomeratic clay), Landslide1 = 1 (landslide near ShangQiaotou), Landslide2 = 1 (Landslide near XiangDuo), Landslide3 = 1 (landslide near JuDian), etc.

Given the nature of the Apriori algorithm, to analyze the association rules of landslide monitoring data, the association rules should be first formulated according to the algorithm logic. In this paper, any elements in the former (X) were randomly combined to formulate a series of association rules (see Table 2). The minimum support was set to 0.01 [29], and the confidence degree of each of the mentioned association rules should be calculated by the algorithm.

Table 2. Description of some association rules (minsupp = 1%).

Rule	Information
Rule 1	If Under_max and Rain_max and River_min and LS1-1 and LandslideD_1 then Landslide1 = 1
Rule 2	If Under_max and Rain_max and River_min and LS1-1 and LandslideD_2 then Landslide1 = 1
Rule 3	If Rain_min and River_min and LS1-0 and LandslideD_3 then Landslide1 = 1
Rule 4	If Under_max and Rain_max and River_max and LS2-1 and LandslideD_3 then Landslide2 = 1
Rule 5	If Under_max and Rain_max and River_min and LS2-1 and LandslideD_3 then Landslide2 = 1
Rule 6	If Under_max and Rain_max and River_min and LS3-1 and LandslideD_2 then Landslide3 = 1
Rule 7	If Rain_max and River_min and LS1-1 and LS3-1 and LandslideD_2 then Landslide3 = 1

On the whole, the association rule exhibiting a confidence above 0.7 is considered a strong association rule [17,29,30]. Association rules with low confidence retain rules with a confidence level above 0.7 are deleted, and a series of useful information is calculated, laying a basis for landslide warning.

The value of confidence in association rules is critical to model prediction. The higher the confidence, the more accurate the landslide prediction will be [29]. Thus, under the same experimental conditions, the confidence of IAprioriMR algorithm and FP-Growth algorithm were compared, and the IAprioriMR algorithm was analyzed in mid-long-term early warning of landslides. According to the experimental results, under the same conditions, the IAprioriMR algorithm (see Table 3) of the same rule exhibits a higher confidence than FP-Growth algorithm (see Table 4), so IAprioriMR can achieve more accurate landslide prediction.

Table 3. Pre-judgment results by IAprioriMR.

NO.	Prejudgment	Confidence	Phenomenon
1	Rain_max and LS1-1 and LanslideD_2	73%	Lanslide1 = 1
2	Rain_max and River_max and LS1-1 and LanslideD_2	85%	Lanslide1 = 1
3	Rain_max and River_min and Under_max and LS2-1 and LanslideD_2	90%	Lanslide2 = 1
4	Rain_max and River_max and Under_max and LS1-1 and LS3-1 and LanslideD_2	95%	Lanslide1 = 1 Lanslide3 = 1

Table 4. Pre-judgment results by FP-Growth.

NO.	Prejudgment	Confidence	Phenomenon
1	Rain_max and LS1-1 and LanslideD_2	70%	Lanslide1 = 1
2	Rain_max and River_max and LS1-1 and LanslideD_2	81%	Lanslide1 = 1
3	Rain_max and River_min and Under_max and LS2-1 and LanslideD_2	83%	Lanslide2 = 1
4	Rain_max and River_max and Under_max and LS1-1 and LS3-1 and LanslideD_2	90%	Lanslide1 = 1 Lanslide3 = 1

To verify the accuracy of the experimental results, test data were taken for analysis, and 21 landslide events were recorded. By calculating the rule with a confidence above 0.7, the pattern successfully determines 16 landslide accidents based on the LS3-1 LS2-1 (see Table 5).

Table 5. Accuracy evaluation of landslide comprehensive prediction.

No.	Input Datasets	Number of Predictions
1	LS3-1 LS2-1	16
2	LS3-1 LS1-1	15
3	LS3-1 LS2-1 LS1-1	15

The IAprioriMR algorithm successfully determines 16 landslide accidents between 2012 and 2013. Thus it is implied that the IAprioriMR algorithm proposed in this paper is feasible in landslide prediction research, and it can technically support the prevention of landslide disasters.

5. Conclusions

During the mid-long-term early warning of landslides, the landslide monitoring data continuously increases over time. The efficiency of the prediction model significantly affects disaster prevention and control. To solve the Apriori algorithm, numerous frequent items are generated when the algorithm runs in the MapReduce framework. The set problem, by simplifying frequent item operations, proposes

the IAprioriMR algorithm. Performance test experiments reveal that the efficiency of the IAprioriMR algorithm is enhanced with the rise in data volume in terms of computational efficiency. Under the same conditions, the IAprioriMR algorithm exhibits a higher confidence than FP-Growth algorithm, capable of achieving more accurate landslide disaster prediction. Thus, it is implied that IAprioriMR algorithm has promising applications and a bright prospect of further development.

In the future work, we will consider comparing the performances of the Apriori algorithm with classical intensity-duration (ID) schemes in a case study, in order to check the significant (or insignificant) improvements for landslide prediction in early warning systems.

Author Contributions: Formal analysis: W.G., X.Z.; Methodology: W.G., X.Z., J.Y., and B.Z.; Writing—review & editing: W.G., X.Z., J.Y., and B.Z.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, Y.; Zhou, L.; Deng, H.; Nan, X. Prediction of landslide displacement using EMD-PSO-ELM with multiple factors. In Proceedings of the International Conference on Digital Home, Guilin, China, 2–6 December 2018; pp. 230–235.
2. Fell, R.; Corominas, J.; Bonnard, C. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng. Geol.* **2007**, *102*, 85–98. [\[CrossRef\]](#)
3. Althuwaynee, O.F.; Pradhan, B.; Park, H.J.; Lee, J.H. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides* **2014**, *11*, 1063–1078. [\[CrossRef\]](#)
4. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [\[CrossRef\]](#)
5. De Luca, L.D.; Versace, P. A comprehensive framework for empirical modeling of landslides induced by rainfall: The Generalized FLAIR Model (GFM). *Landslides* **2017**, *14*, 1009–1030. [\[CrossRef\]](#)
6. Ermini, L.; Catani, F.; Casagli, N. Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* **2005**, *66*, 327–343. [\[CrossRef\]](#)
7. Wang, X.M.; Niu, R.Q. Landslide intelligent prediction using object-oriented method. *Soil Dyn. Earthq. Eng.* **2010**, *30*, 1478–1486. [\[CrossRef\]](#)
8. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir. *China Eng. Geol.* **2016**, *204*, 108–120. [\[CrossRef\]](#)
9. Wang, G.; Sassa, K. Factors affecting rainfall-induced flowslides in laboratory flume tests. *Géotechnique* **2001**, *51*, 587–599. [\[CrossRef\]](#)
10. Jin, X.; Zeng, J.; Liu, X. Application of GM (1,1) optimized model in prediction of landslide. In Proceedings of the International Conference on Natural Computation, Haikou, China, 24–27 August 2007; pp. 735–739.
11. Xie, P.; Zhou, A.; Chai, B. The application of long short-term memory (LSTM) method on displacement prediction of multifactor-induced landslides. *IEEE Access* **2019**, *7*, 54305–54311. [\[CrossRef\]](#)
12. Hossein, S.M.; Masoud, M.; Himan, S. Big data in geohazard; pattern mining and large scale analysis of landslides in Iran. *Earth Sci. Inform.* **2019**, *12*, 1–17.
13. Mandavilli, V.S.; Madamala, N. Detection of natural disaster affected areas using R. *Int. J. Inf. Technol.* **2018**, *10*, 123–131. [\[CrossRef\]](#)
14. Ma, J.; Tang, H.; Hu, X. Identification of causal factors for the Majiagou landslide using modern data mining methods. *Landslides* **2017**, *14*, 311–322. [\[CrossRef\]](#)
15. Webb, G.I. Discovering significant patterns. *Mach. Learn.* **2007**, *68*, 1–33. [\[CrossRef\]](#)
16. Liao, C.W.; Perng, Y.H. Data mining for occupational injuries in the Taiwan construction industry. *Saf. Sci.* **2008**, *46*, 1091–1102. [\[CrossRef\]](#)
17. Wu, X.; Zhan, F.B.; Zhang, K. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. *Environ. Earth Sci.* **2016**, *75*, 146–162. [\[CrossRef\]](#)

18. Wei, L.; Guangming, L.; Yachao, S.; Junlong, L.; You, Z. Optimization and application in medical big document-data of Apriori algorithm based on MapReduce. In Proceedings of the International Conference on Computer Communication and Informatics, Coimbatore, India, 7–9 January 2016; pp. 1–5.
19. Belhassena, A.; Wang, H. Trajectory big data processing based on frequent activity. *Tsinghua Sci. Technol.* **2019**, *24*, 317–332. [[CrossRef](#)]
20. Duaimi, M.G.; Salman, A. Association rules mining for incremental database. *Int. J. Adv. Res. Comput. Sci. Technol.* **2014**, *2*, 346–352.
21. Luna, J.M.; Cano, A.; Pechenizkiy, M.; Ventura, S. Speeding-up association rule mining with inverted index compression. *IEEE Trans. Cybern.* **2016**, *46*, 3059–3072. [[CrossRef](#)]
22. Luna, J.M.; Padillo, F.; Pechenizkiy, M. Apriori versions based on mapreduce for mining frequent patterns on big data. *IEEE Trans. Cybern.* **2018**, *48*, 2851–2865. [[CrossRef](#)]
23. Zhao, Y.; Wu, J.; Liu, C. Dache: A data aware caching for big-data applications using the MapReduce framework. *Tsinghua Sci. Technol.* **2014**, *19*, 39–50. [[CrossRef](#)]
24. Wang, S.Q.; Yang, Y.-B.; Chen, G.-P.; Gao, Y.; Zhang, Y. Mapreduce based closed frequent itemset mining with efficient redundancy filtering. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium, 10–13 December 2012; pp. 449–453.
25. Ding, M.; Su, P.; Wang, H. RETRACTED ARTICLE: Distribution characteristics of debris flows and landslides in Three Parallel Rivers Area. *Nat. Hazards* **2012**, *61*, 879. [[CrossRef](#)]
26. Yi, S.; Ning, W.; Peng, L. Changes in livestock migration patterns in a Tibetan-style agropastoral system: A study in the three-parallel-rivers region of Yunnan, China. *Mt. Res. Dev.* **2007**, *27*, 138–146.
27. Matsuura, S.; Asano, S.; Okamoto, T. Relationship between rain and/or meltwater, pore-water pressure and displacement of a reactivated landslide. *Eng. Geol.* **2008**, *101*, 49–59. [[CrossRef](#)]
28. Montella, A.; Aria, M.; D'Ambrosio, A.; Mauriello, F. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accid. Anal. Prev.* **2012**, *49*, 58–72. [[CrossRef](#)] [[PubMed](#)]
29. Fang, G.; Xiong, J.; Tu, C.; Luo, A. An algorithm of mining spatial topology association rules based on Apriori. In Proceedings of the International Conference on Intelligent Computing and Intelligent Systems, Xiamen, China, 29–31 October 2010; pp. 101–104.
30. Jencso, K.G.; McGlynn, B.L. Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resour. Res.* **2011**, *47*, 431–435. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).