Study on Temperature Related Factors of Asphalt Pavement Based on APRIORI

W Y Lin[1], Z H Zhang[2], X Qiu[3], W Y Xu[3] and H J Hong[3]

1 Jinhua Traffic Engineering Supervision Station, Jinhua, 321013, China
2 Zhejiang Zhengda Testing Technology Co., Ltd., Jinhua, 321000, China
3 Road and Traffic Engineering Research Center, Zhejiang Normal University, Jinhua, 321004, China
E-mail: zjnuhhj@outlook.com

Abstract. Asphalt is a typical temperature-sensitive material. In order to explore the correlation between surface temperature of asphalt pavements and meteorological factors, the association rules mining algorithm (Apriori) was applied to identify the key factors affecting the surface temperature of asphalt pavement. Firstly, errors and missing data in the meteorological dataset were cleaned. Then, Apriori was applied to identify the key factors affecting the asphalt pavement temperature. The results indicate that Apriori would perform an excellent ability to analyze the correlation rules between meteorological factors with the minimum confidence of 0.8 and the key meteorological factors which affect the temperature change of asphalt pavements are excavated including air temperature, air pressure, dew point temperature and relative humidity. The research would serve as a technical support for the machine learning algorithms applied in the field of the association rule analysis.

1. Introduction

Asphalt is a common material which was widely used in pavement due to its properties, but asphalt is a typical temperature-sensitive material, and the change of road surface temperature and state will directly affect the road traffic safety level and traffic efficiency, so meteorological conditions are important factors affecting road surface temperature change. Therefore, how to accurately identify the key meteorological factors affecting the pavement temperature has important theoretical and practical significance for road traffic safety.

Now, researchers have conducted extensive research on the variation law of road surface temperature. In 1975, Barder [1] first used the heat conduction equation to establish a model for the highest temperature prediction of the road surface and performed statistical analysis. With the gradual improvement of the prediction model of asphalt pavement temperature field, researchers in various countries began to pay attention to the selection of input parameters of the model. Hermansson [2] and Grimbacher [3] selected the appropriate meteorological factors to establish the model to predict the temperature field of the road surface. As the amount of data increases, the process of model establishment becomes complicated, also the precision of the prediction model is hard to be guaranteed. In addition, the input parameters of the model set by artificial can reduce the prediction accuracy. Therefore, how to accurately identify the relevant factors is the research emphasis.

The emergence of machine learning method has broken the limitations of traditional methods. Also, its excellent performance has been shown in association rule mining and model ensemble prediction. Association rules mining is an unsupervised learning algorithm for machine learning, and Apriori is its representative algorithms which have been proposed by Agrawal [4] in 1994 for the first time. Apriori

is a frequent itemsets mining algorithm for extracting association rules using frequent itemsets obtained by loop iteration. Apriori algorithm was originally used for market basket analysis. Then numerous scholars such as Borgelt [5] began to apply it to the analysis of association rules of data. The result demonstrates that Apriori has a greater advantage than Eclat in mining the closed itemsets association rules. Guo [6] also found that association rules drew by Apriori have powerful ability in handling the forecasted wind speed values correction. However, with the exponential increase in data volume, low efficiency of Apriori operations becomes prominent, thus the improved form of Apriori algorithm has developed rapidly. For example, in order to improve the efficiency of association rule mining, Park [7] reduced the size of candidate itemsets by using an appropriate hash function. Similarly, by improving the operation process of Apriori algorithm, Han [8] proposed a method called FP-growth (frequent pattern growth) that succeeded in eliminating candidate generation. As can be seen from the foregoing, Apriori algorithm is widely used in association rule mining and has a good application effect in the field of association analysis. Therefore, in this paper, hash-based Apriori algorithm is used to mine the correlation between meteorological factors and asphalt pavement temperature.

This paper, in the light of related research, selects the winter meteorological observation data from December 2015 to January 2018 in Jinhua area of Zhejiang Province of China. The research goal is to identify key meteorological factors affecting the road temperature based on the association rules mined by Apriori algorithm. The study would serve as a technical support for the accurate prediction of winter asphalt pavement temperature.

2. Data Preprocessing

2.1. Data Cleaning
Data cleaning is intended to handle anomalous data in data sets, standardize dataset forms, and provide numerically accurate, formal-standard datasets for data mining. In this study, python platform was used to organize the data, and 3261 sets of valid data are generated.

2.2. Data Discretization
The meteorological data used for association analysis mainly include: air temperature (T), dew point temperature (D), air pressure (P), relative humidity (H), wind speed (W), precipitation (R), visibility (V) and asphalt pavement temperature (APT). Before the Apriori algorithm analyzes the correlation of meteorological data, it is necessary to discretize each data [6]. And the discretization of meteorological data is usually based on national grading standards of meteorological data, and the degree of grading of each factor is reduced by enlarging the grading interval to improve the operational efficiency of the Apriori algorithm. In this paper, the Short-term Weather Forecast (GB/T 21984-2008) [9] is used as the basis for the classification of meteorological factors, and the classification of meteorological factors is given in Table 1.

Table 1. Classification of Meteorological Factors.

Factors	Level 1	Level 2	Level 3	Level 4
T (℃)	T≤0	0<T≤13.9	13.9<T≤21.9	21.9<T
D (℃)	D≤-9.9	-9.9<D≤0	0<D≤9.9	9.9<D
P (Hpa)	P≤1010	1010<P≤1020	1020<P≤1030	1030<P
H (%)	H≤25	25<H≤50	50<H≤75	75<H
W (m/s)	W≤3.4	3.4<W≤10.7	10.7<W≤20.7	20.7<W
R (mm/h)	R≤2.6	2.6<R≤8.1	8.1<R≤16	16<R
V (m)	V≤50	50<V≤500	500<V≤2000	2000<V
APT (℃)	APT≤0	0<APT≤10	10<APT≤20	20<APT

Data in the meteorological database are subjected to value conversion according to the scale quantization standard and get the set of transaction dataset for association rule mining. The level of 8 meteorological factors within an hour is taken as an itemset I, which can be expressed as I = {Ti, Di,

Pi, Hi, Wi, Ri, Vi, APTi}, where i is the level of each meteorological factor. For example, the data item at 8:00 on December 5, 2015 is {6.4, 1.9, 1020.9, 73, 2, 0.2, 94, 8.1}, and each data in item corresponds to {temperature, dew point temperature, air pressure, relative humidity, wind speed, Precipitation, visibility, asphalt pavement temperature}, Referring to the meteorological factor classification of Table 1, the data items can be converted into itemset as I = {T2, D3, P3, H3, W1, R1, V2, APT2}. According to the conversion rule, all the data items in the database can be converted into an itemset, and then the transaction set for association rule mining can be obtained.

3. Association Rules Mining

3.1. Association Rule Mining Process
Apriori is a two-stage recursive algorithm. The first step is to find frequent itemsets from the transaction dataset that satisfy the minimum support, and the second step is to derive association rules from frequent items that satisfy the minimum confidence [5]. Therefore, it is necessary to preset minimum support and minimum confidence before the association rules are mined. The equation for support and confidence is as follows:

$$support\left(A \rightarrow B\right) = P(A \cup B)$$
$$confident\left(A \rightarrow B\right) = P(A \cup B) \, / \, P\left(A\right)$$

(1)

Among them, support indicates the probability that A and B appear simultaneously, and confidence indicates the probability that A and B appear simultaneously on the premise of A has occurred.

Apriori algorithm has two steps in the process of finding frequent itemsets: Join step and Prune step. In the join step, Apriori scans the database and searches for $K+1$ itemset depend on K itemset, finally extract all the frequent itemsets that satisfy the minimum support. The prune step is interleaved with the join step, and the purpose is to discard the itemset that does not satisfy the minimum support and retain the frequent itemset.

After all frequent itemsets are found by Apriori algorithm, derived association rules from the frequent itemset, discard all association rules that do not meet the minimum confidence, and the retained association rule is the mining result of the Apriori algorithm.

3.2. Association Rule Mining Results
The weather with a road surface temperature below zero is rarely in Jinhua area of Zhejiang Province. For the sake of the comprehensiveness and reliability of the association rule mining results, it is necessary to pay attention to the association rules of asphalt pavement temperature below zero. Therefore, the Apriori algorithm is used to perform a large number of trials on the transaction set to determine the minimum support and minimum confidence. In association rule mining analysis, in order to make the extracted association rules have higher accuracy, the minimum confidence value is generally greater than 0.75. So, the minimum confidence is set to 0.8 in this study. On this foundation, the trial results are presented in Figure 1. It shows that the two-dimensional association rule with zero asphalt pavement temperature is stable to three when the support degree is no more than 0.003. Therefore, the minimum support is set to 0.003.
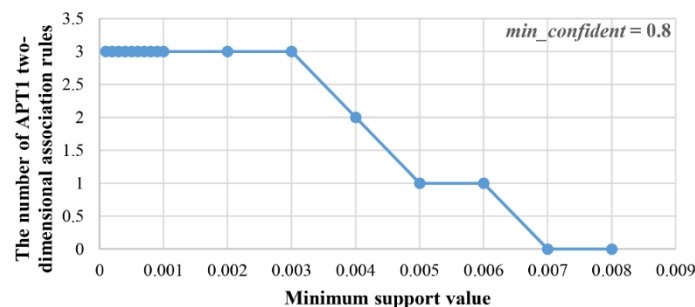

Figure 1. Apriori algorithm trial results.

Apriori algorithm mines 7 association rules for APT1; 28 association rules for APT2; 22 association rules for APT3; and due to the fact that the asphalt pavement temperature in the APT4 grade is rarely in winter, so only 1 two-dimensional association rule is obtained for the mining result. By analyzing the relationship between the association rules of each dimension, it is found that the Three-dimensional association rules can cover all meteorological factors, and a large number of invalid association rules appear from the four-dimensional association rules. Therefore, the analysis focuses on the analysis of one, two and three-dimensional association rules. It is worth mentioning that the existence of invalid association rules will directly affect the accuracy of the analysis results and should be eliminated in the analysis process. For example, in the analysis of the three-dimensional association rule, if the confidence of the two-dimensional association rule {H3, P4}→APT1 is 0.8333, and also the confidence of the three-dimensional association rule {H3, W1, P4}→APT1 is 0.8333, then the three-dimensional association rule is determined to be an invalid association rule, and needs to be discarded.

3.3. Association Rule Analysis
Figure 2 to Figure 4 show the statistical analysis results of the two-dimensional association rules. The analysis results show that in the two-dimensional association rules of different asphalt pavement temperature grades, two meteorological factors, dew point temperature (D) and air pressure (P), appeared in antecedent, and both grades change with the change of asphalt pavement temperature grade. It can be known that the dew point temperature and air pressure are 2 key factors affecting the temperature change of asphalt pavement.

Figure 5 to Figure 7 show the statistical analysis results of the three-dimensional association rules. The analysis results are demonstrated that the two factors of relative humidity (H) and air temperature (T) appear in different asphalt pavement temperature grade correlation rules, and all change with the pavement temperature grade. The result is indicating that relative humidity and temperature are also two key factors affecting asphalt pavement temperature change. In addition, there is no similar pattern of wind speed (W), visibility (V) and precipitation (R).
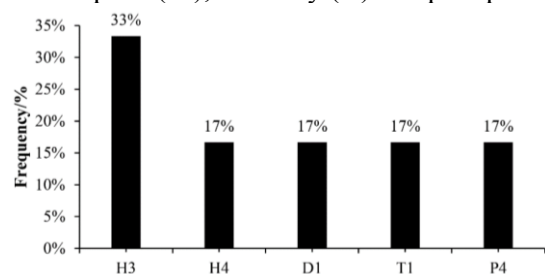

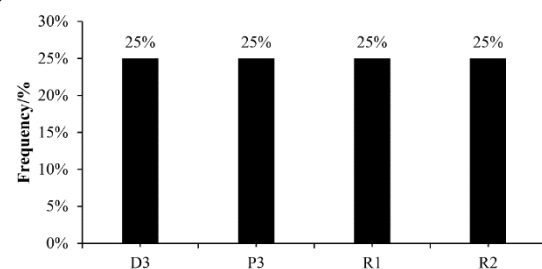**Figure 2.** Statistics frequency for APT1 two-dimensional association rules..

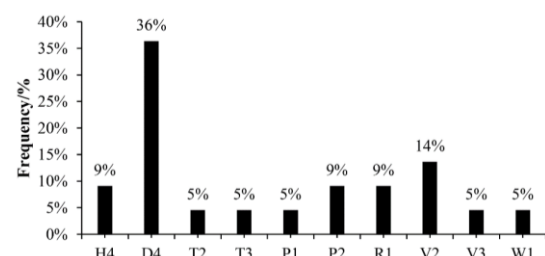**Figure 3.** Statistics frequency for APT2 two-dimensional association rules.

Figure 4. Statistics frequency for APT3 two-dimensional association rules.
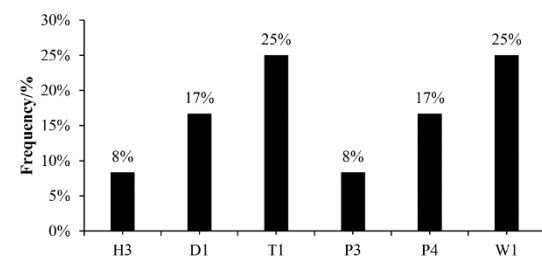

Figure 5. Statistics frequency for APT1 three-dimensional association rules.
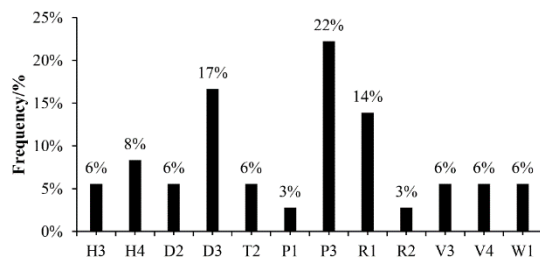
Figure 6. Statistics frequency for APT2 three-dimensional association rules.
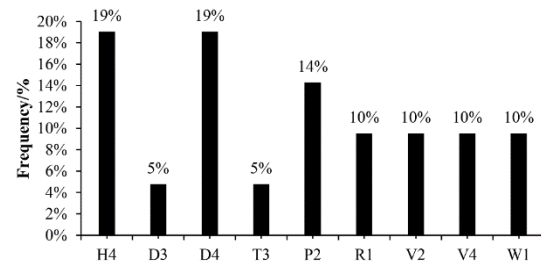

Figure 7. Statistics frequency for APT3 three-dimensional association rules.

According to the analysis results of the above association rules, there are four fundamental meteorological factors affecting the temperature change of asphalt pavement: dew point temperature, air pressure, relative humidity and air temperature.

4. Conclusion

The following major findings were achieved from this study:

(1) Apriori algorithm preforms well in terms of mining association rules, and compared with the linear fitting methods, it has the advantage of obtaining the correlation between different factors without drawing figures or calculating the correlation coefficient, which greatly improves the efficiency of the correlation analysis, especially dealing with big data.

(2) According to the mining results of Apriori algorithm, there are four key meteorological factors affecting the temperature change of asphalt pavement in Jinhua area of central Zhejiang Province: dew point temperature, air pressure, relative humidity and temperature. Also, the meteorological factors collected in this paper are relatively small, so in the further study the number of meteorological factors can be further increased.

References

[1]   Barber E S 1957 Calculation of Maximum Pavement Temperatures from Weather Reports Highway Research Board Bulletin

[2]   Hermansson 2001 Mathematical Model for Calculation of Pavement Temperatures: Comparison of Calculated and Measured Temperatures *Transportation Research Record Journal of the Transportation Research Board*

[3]   Grimbacher T and Schmid W 2005 Nowcasting Precipitation, Clouds, and Surface State in Winter. *Atmospheric Research*,. **77** pp 378-387

[4]   Agarwal R and Srikant 1994 R Fast Algorithms for Mining Association Rules. P*roc. of the 20th VLDB Conference* p 487-499

[5]   Borgelt C 2003 Efficient Implementations of Apriori and Eclat *Proc.ieee Icdm Workshop on Frequent Item Set Mining Implementations. ceur* Workshop Proceedings **90**

[6]   Guo Z, Chi D and Wu J 2014 A New Wind Speed Forecasting Strategy Based on the Chaotic Time Series Modelling Technique and the Apriori Algorithm *Energy Conversion & Management* **84** pp 140-151

[7]   Park J S, Chen M S and Yu P S 1995 An Effective Hash-based Algorithm for Mining Association Rules *ACM* p 175-186

[8]   Han J, Pei J and Yin Y 2000 Mining Frequent Patterns Without Candidate Generation *ACM SIGMOD International Conference on Management of Data* p 1-12

[9]   GB/T 21984-2008. Short-term Weather Prediction.