Development Trend of Intelligent Transportation in the Era of Big Data

Jing Wang

College of Intelligent Manufacturing and Automobil, Chongqing Vocational College of Transportation, Chongqing, China

*Corresponding author e-mail: zfshi178@bjtu.edu.cn

Abstract. The arrival of the big data era has brought great convenience to people's real life. All aspects of our lives follow the development of the big data era. In the context of the era of big data, through data collection and analysis, the current urban transportation has been reasonably developed, laying a good foundation for the development of intelligent transportation in the future, and can solve various current problems. Based on the analysis of the traffic system, this article finds that the real-time road condition prediction of common functions is highly correlated with route guidance. Therefore, this paper uses a method based on historical data to predict road conditions, and uses the prediction results to dynamically allocate road segments, and then uses the relevant single-source shortest path algorithm to simulate the optimal path. By analyzing the application advantages of data mining in intelligent transportation, Baum Welch's algorithm is used for data mining in a real-time traffic prediction system based on historical data. Real-time road conditions are matched with historical data models to predict road conditions. Finally, analyzed and designed the big data cloud platform framework applied to it to achieve higher efficiency. The experimental research results show that, by studying the development of urban intelligent transportation systems, the purpose of this paper is to make better use of existing resources to process data more efficiently, and then combine distributed data mining algorithms to optimize the processed data and apply it to in the intelligent transportation system, an intelligent management and control system that is more efficient, accurate, and user experience better can be established.

Keywords: Big Data Era; Intelligent Transportation; Baum-Welch Algorithm; Intelligent Control System

1. Introduction

In the application and development of big data technology in urban intelligent transportation, the concept and industry of big data in China started late [1]. With the rapid development of economy and science and technology at this stage, it will exceed 22 billion yuan in the next three years. With the attention and development of big data in all walks of life, big data has been applied to all walks of life, including transportation, medical, biotechnology, retail, agricultural production and personal services, and new services and technologies related to big data[2-3]. Modern intelligent transportation system
mainly includes data perception part, software application part, data analysis and prediction part and optimization management part. The body perception part is to extract timely and effective data of current traffic conditions. The software application part is the comprehensive processing and data analysis of the perception data, and timely early warning measures are taken according to the actual situation [4].

Strengthen the in-depth research of intelligent transportation in the era of artificial intelligence with big data technology. This technology is conducive to improving the development level of urban transportation and improving the traditional traffic congestion environment [5]. Promote the technology and advantages of its intelligent system network in the process of operation more obvious, fully and effectively deal with road traffic accidents. The quality can be gradually improved, deepening the degree of intelligence in the application process. At the same time, it is necessary to improve the level of awareness of the development of intelligent transportation, so that its network performance can be continuously optimized to meet the development requirements of the era of big data artificial intelligence [6]. The processing of data is the development of the future intelligent model and the core of the technology for the sustainable development of intelligent transportation in the future. The realization of the original intelligent transportation requires people to model the big data after research. Determine the calculation criteria and process the data through the results [7].

In addition, because people cannot make pre-judgments about the existing transmission parameters, the artificial intelligence model based on big data cannot realize the transmission between systems. The processing of big data is the core of the optimized structure of the industry, and the development of intelligent transportation should be structure-oriented. At this stage, due to the low level of data processing in the overall industry, there are certain difficulties in the collection and sorting of big data, and the development of the intelligent transportation industry is unevenly distributed. The underdeveloped cities in the west do not have basic information processing functions. For data collection The application of the system and the system are very difficult [8-9]. However, in areas where data is built earlier, due to the lack of certain standards for data processing, there are differences in the construction and application of data systems, and the transportation department and the management department are independent. Moreover, there are many systems for data construction and large technical differences, leading to a wide variety of sources of big data, and the problem of data independence is very serious [10].

2. Method

2.1. Baum-Welch algorithm
The Baum-Welch algorithm is a supervised model learning and training algorithm. It is assumed that a given training data value contains S observation sequences of length \{O₁,O₂,..,Oₘ\} without a corresponding state sequence, So the formula is as follows:

\[
P(O,λ) = \sum_1 P(O|I,λ)P(I,λ)
\]  

(1)

Its parameter learning can be realized by EM algorithm.
Step E of the EM algorithm: Find the Q function \(Q(λ,κ)\)

\[
Q(λ,κ) = \sum LOG P(O|I,λ)P(I,λ)
\]  

(2)

Then the function \(Q(λ,κ)\) can be summed with:

\[
Q(λ,κ) = \sum \sum \log P(O|I,λ)P(I,λ) + \sum \sum \log a_{i|t+1} \sum P(O|I,λ) + \sum \sum \log b_{t|O} \sum P(O|I,λ)
\]  

(3)
M step of the EM algorithm: Maximize the Q function \(Q(\lambda, \lambda)\) to find the model parameters \(\lambda=(A, B, \pi)\).

Finally, combining the above-mentioned formula, the formula that can be obtained is as follows:

\[
b_j(k) = \frac{\sum_{t=1}^{T} p(O_{it} = j) l(O_{it} = v_j)}{\sum_{i=1}^{T} p(O_{it} = j) l}\tag{4}
\]

Since there are many similar data sets in a given sample, and the training parameter model is based on a single sample. For the first simulation, two solutions were proposed. First, the mean training method is used for the data fusion of the model input, that is, the observation data of the sample.

2.2. Intelligent Road Condition Assessment Model

In practical applications, the overall distribution \(P\) is usually unknown, so it is impossible to directly obtain the conditional expectations of the impact of road conditions to evaluate an accurate prediction plan, so it is necessary to estimate it. Considering that the effect of the application is related to the application plan and the corresponding traffic characteristics and behavior performance, based on the regression model, the evaluation model of the accurate road condition plan is constructed as follows:

\[
E(Y | A, X) = y^T \bar{X} + A(\beta^T \bar{X})\tag{5}
\]

\[
\beta^T \bar{X} = E(Y | A = 1, X) - E(Y | A = 0, X)\tag{6}
\]

In summary, the formula for parameter estimation of the intelligent road condition assessment model is:

\[
\beta_{\text{LASSO}} = \min_{\beta} \{ ||y - X\beta||^2 + \gamma \sum_{j=1}^{p} |\beta_j| \} \tag{7}
\]

Due to the singularity of the penalty term at the zero point, the Lasso method can make the partial coefficient estimation of the model accurate to zero. Based on this, the coefficients of some unimportant variables (variables with smaller parameter estimates) are likely to be compressed to zero. For important variables with large parameter estimates, the Lasso method can slightly compress the coefficients to ensure the accuracy of parameter estimates.

2.3. Establish an Intelligent Traffic Safety Management System

In addition to the above formula, in the process of modern intelligent transportation construction, can the construction and development level of the data security management system be truly guaranteed? It is necessary to strengthen the management of the confidentiality of information and data and the boundaries of the information and data disclosure party, and establish a more complete and safe traffic safety protection and transportation system guarantee system. On this basis, ID management technology and system security technology have been added to truly improve the information security management level of intelligent transportation. In intelligent transportation systems, risks are also classified according to the characteristics of diverse data types. For different risk levels, corresponding protection and treatment strategies are proposed, which can prevent waste of resources or poor management, and propose the best problem-solving strategies.

3. Experiment

3.1. Safety Rating Model

The fitting effect of the rating model is expressed by the pseudo R-square statistic, and its meaning is a proportional model explained by the self-edited amount of the total variation times model of the dependent variable. It can be seen that:
Cox&Snell \[ R^2 = 1 - \left[ \frac{\text{Ln}(L_0)}{\text{Ln}(L)} \right]^2 = 1 - e^{\left\{ \frac{\text{Cox&Snell} - R^2}{2} \right\}} \]

The coverage ratio of the result is the actual rating sample, the correct rating proportion model is as follows:

\[ \text{RESF} = \frac{T_A}{T_A + FA} \]

The accuracy of the result is the actual rating sample, the correct rating proportion model is as follows:

\[ \text{RESJ} = \frac{T_A}{T_A + FN} \]

According to 50 synthesizable feature dimensions, in order to avoid information errors and omissions and increase the complexity of model calculation, it is necessary to expand feature selection to generate a relatively small and optimal feature set. In this paper, \( S \) is the main feature selection method. The \( S \) value is the value index of information, which reflects the degree of influence of the characteristic value on the intelligent transportation risk rating. The larger the value obtained, the stronger the predictive ability of the target:

\[ \text{WOE}_i = \ln \left( \frac{p_{y_i}}{p_{n_i}} \right) \]

\[ S = \sum (p_{y_i} - p_{n_i}) \cdot \text{WOE}_i \]

The model describes the regression results between big data storage ratings and related indicators. The results show that as the computing performance of the computer's processor carried by the data cloud increases, the larger the data stream it carries, the greater the ability to withstand sudden consequences, and then use multiple algorithms to intelligently Optimize operation and maintenance management, and then get the optimal solution.

3.2. Design of Experimental Research

Big data technology is used to analyze and process the accumulated historical data, and the prediction module of intelligent transportation system is used to verify the prediction model. The experimental data uses the previous road condition vector and the previous 7 road condition states of a main road in a certain city in East China as the input comparison time of the algorithm. Collect a moment for matching. When the current time is 8 o'clock in the morning, the collected real-time measured average vehicle speed is 45km/h. It is necessary to predict the average vehicle speed at the next time 9.05, then the historical database is close to All dates with an average speed of 38km/h~50km/h at nine o'clock in a year are all extracted, and the average speed of these dates at 10.05 is used as the Baum-Welch algorithm, and the maximum classification is the predicted speed of the day.

Collect five moments for matching: Also suppose it is now ten o'clock, and you need to predict the average speed of the next ten and five minutes, then you need to find the first five moments in the real-time database (for example, nine thirty, nine o'clock) 35, 9.40, 9.45, and 9.50 average vehicle speed, and then extract all the dates in the historical database that are closest to the average vehicle speed of these five times in the past year, and use The average vehicle speed at 10:05 on these dates is the Baum-Welch algorithm, and the maximum classification is the predicted vehicle speed for the day.
Perform the Baum-Welch algorithm for the road conditions at the previous moment and the data within one year of real-time traffic conditions and the Baum-Welch algorithm for the traffic data at the first five moments of the real-time traffic conditions and the data within one year. The running speed of the interval is used as the predicted road condition.

4. Results

4.1. Analysis of Experimental Test Results

Table 1. Comparison and analysis table of real-time road conditions and predicted road conditions

| Acquisition time | Actual vehicle speed (Km/h) | Speed at a certain moment (Km/h) | Prediction of vehicle speed at six moments (Km/h) |
|------------------|----------------------------|---------------------------------|-----------------------------------------------|
| 9:30             | 37                         | 35                              | 53                                            |
| 9:35             | 47                         | 39                              | 47                                            |
| 9:40             | 53                         | 56                              | 41                                            |
| 9:45             | 72                         | 54                              | 39                                            |
| 9:50             | 55                         | 47                              | 40                                            |

Figure 1. Comparison of real-time road conditions and predicted road conditions

From the comparison of the data in Table 1 and Figure 1, from the experimental comparison results of the road condition prediction and the actual road condition, the prediction result is related to the use of historical samples. The road condition at the previous moment and the data within one month of real-time road conditions are collected for Baum- The accuracy of the Baum-Welch algorithm is different between the Welch algorithm and the road condition data collected at the first 8 moments of real-time traffic conditions and the data within three months. Based on the historical data real-time road condition prediction model, the average absolute error of the Baum-Welch algorithm is less than 8km/h, and the average absolute error of the error is about 6.85km/h; and the real-time road condition is collected. The forecast accuracy of the Baum-Welch algorithm for the road condition data at the first three moments and the data within 3 months is relatively high. The average absolute error of the average speed of the predicted road conditions and the average speed of the actual road conditions is less than 6km/h, and the average The relative error is basically no more than 2%, and the mean square error of the error is about 1.67km/h. Therefore, as the time of collecting road conditions increases and the time for comparison in the historical database increases, the accuracy of the system prediction can be increased, and the second method is adopted. The degree of dispersion of the error between the predicted result and the actual result is different.
Figure 2. The subsequent state distribution of the Baum-welch algorithm for processing real-time road conditions

The data shown in Figure 2, the intelligentization of data cloud traffic management, and the processing of big data technology and artificial intelligence technology, will lead to the possibility of data interruption, and the consequences of this situation are more serious. On the other hand, in a large number of normal analysis conditions and abnormal analysis operation conditions, the intelligent traffic management system mainly focuses on the abnormal analysis conditions, and the rest of the normal operation status is actually out of your consideration. Therefore, this article passed Design experiments, perform supervised learning and semi-supervised learning on the same data set with the same labeled part, use consistent evaluation indicators to evaluate the performance of the model, and finally select the best performing model to evaluate the status of the intelligent risk rating of the data cloud Make predictions.

5. Conclusion
To sum up, the construction of intelligent transportation systems in the era of big data and artificial intelligence is actually a very good system. We should follow the development trend of the current era, make reasonable use of big data and artificial intelligence, and apply them to us. In all aspects of life, one of the very important aspects is our urban traffic problem. We should make reasonable use of it to improve our current urban traffic dilemma and work to better provide urban traffic services. Specific research ideas the main thing is that we should focus on improving the integrity and effectiveness of database information, regularly testing the efficiency of traffic detectors, and focusing on the introduction of intelligent elements through three aspects. The future development of transportation is moving towards data and intelligence. Through in-depth research on big data and artificial intelligence technology, the probability of traffic accidents is reduced, which can make the development of intelligent transportation more rapid.

References
[1] Xu W, Zhou H, Cheng N, et al. Internet of Vehicles in Big Data Era[J]. IEEE/CAA Journal of Automatica Sinica, 2018, 5(1):19-35.
[2] Zheng C, Lei L, Zhenyu Z, et al. Learn to Cache: Machine Learning for Network Edge Caching in the Big Data Era[J]. IEEE Wireless Communications, 2018, 25(3):28-35.
[3] Shen B, Choi T M, Chan H L. Selling green first or not? A Bayesian analysis with service levels and environmental impact considerations in the Big Data Era[J]. Technological Forecasting and Social Change, 2019, 144(JUL.):412-420.
[4] Cheng F, Li H, Brooks B W, et al. Retrospective Risk Assessment of Chemical Mixtures in the Big Data Era: An Alternative Classification Strategy to Integrate Chemical and
Toxicological Data[J]. Environmental science And Technology, 2020, 54(10):5925-5927.

[5] Sandt A , Al-Deek H . An optimization approach for deployment of intelligent transportation systems wrong-way driving countermeasures[J]. Journal of Intelligent Transportation Systems, 2019, 24(2):1-14.

[6] Arghya, Guchhait, B, et al. A hybrid V2V system for collision-free high-speed internet access in intelligent transportation system[J]. European transactions on telecommunications, 2018, 29(3):187-195.

[7] Taimouri A , Emamisaleh K . Providing Performance Evaluation Indicators for Intelligent Transportation Systems (The Case Study of Tehran-Karaj Freeway Located in Iran)[J]. Journal of Transportation Technologies, 2020, 10(2):144-153.

[8] Sassi I , Anter S , Bekkhoucha A . A New Improved Baum-Welch Algorithm for Unsupervised Learning for Continuous-Time HMM Using Spark[J]. International Journal of Intelligent Engineering and Systems, 2020, 13(1):214-226.

[9] Steeneck D , Eng-Larsson F . The Baum–Welch algorithm with limiting distribution constraints[J]. Operations Research Letters, 2018, 46(6):563-567.

[10] Wei Y , Wu Y , Zhang X , et al. Fuzzy Self-Tuning PID-Based Intelligent Control of an Anti-Wave Buoy Data Acquisition Control System[J]. IEEE Access, 2019, PP (99):1-1.