APEM journal
Advances in Production Engineering & Management
Volume 16 | Number 3 | September 2021 | pp 285–296
https://doi.org/10.14743/apem2021.3.400
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Using the gradient boosting decision tree (GBDT) algorithm for a train delay prediction model considering the delay propagation feature

Zhang, Y.D.[a,*], **Liao, L.**[a], **Yu, Q.**[a], **Ma, W.G.**[a], **Li, K.H.**[b]

[a]School of Information Science and Technology, Southwest Jiaotong University, Chengdu, P.R. China
[b]School of management, Xihua University, Chengdu, P.R. China

ABSTRACT

Accurate prediction of train delay is an important basis for the intelligent adjustment of train operation plans. This paper proposes a train delay prediction model that considers the delay propagation feature. The model consists of two parts. The first part is the extraction of delay propagation feature. The best delay classification scheme is determined through the clustering method of delay types for historical data based on the density-based spatial clustering of applications with noise algorithm (DBSCAN), and combining the best delay classification scheme and the k-nearest neighbor (KNN) algorithm to design the classification method of delay type for online data. The delay propagation factor is used to quantify the delay propagation relationship, and on this basis, the horizontal and vertical delay propagation feature are constructed. The second part is the delay prediction, which takes the train operation status feature and delay propagation feature as input feature, and use the gradient boosting decision tree (GBDT) algorithm to complete the prediction. The model was tested and simulated using the actual train operation data, and compared with random forest (RF), support vector regression (SVR) and multilayer perceptron (MLP). The results show that considering the delay propagation feature in the train delay prediction model can further improve the accuracy of train delay prediction. The delay prediction model proposed in this paper can provide a theoretical basis for the intelligentization of railway dispatching, enabling dispatchers to control delays more reasonably, and improve the quality of railway transportation services.

ARTICLE INFO

Keywords:
Train delay prediction;
Actual train operation data;
Delay type identification;
Delay propagation feature extraction;
Density-based spatial clustering of applications with noise (DBSCAN);
k-nearest neighbor (KNN);
Gradient boosting decision tree (GBDT);
Random forest (RF);
Support vector regression (SVR);
Multilayer perceptron (MLP)

**Corresponding author:*
ydzhang@swjtu.edu.cn
(Zhang, Y.D.)

Article history:
Received 24 July 2021
Revised 25 October 2021
Accepted 28 October 2021

References

[1] Huang, P., Peng, Q., Wen, C., Yang, Y. (2018). Random forest prediction model for Wuhan-Guangzhou HSR primary train delays recovery, *Journal of the China Railway Society*, Vol. 40, No. 7, 1-9.
[2] Wen, C., Li, Z., Huang, P., Tian, R., Mou, W., Li, L. (2019). Progress and perspective of data-driven train delay propagation, *China Safety Science Journal*, Vol. 29, No. S2, 1-9, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.001.
[3] Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D. (2018). Train delay prediction systems: A big data analytics perspective, *Big Data Research*, Vol. 11, 54-64, doi: 10.1016/j.bdr.2017.05.002.
[4] Wang, P., Zhang, Q.-P. (2019). Train delay analysis and prediction based on big data fusion, *Transportation Safety and Environment*, Vol. 1, No. 1, 79-88, doi: 10.1093/tse/tdy001.

[5] Shi, R., Xu, X., Li, J., Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization, *Applied Soft Computing*, Vol. 109, Article No. 107538, doi: 10.1016/j.asoc.2021.107538.
[6] Nair, R., Hoang, T.L., Laumanns, M., Chen, B., Cogill, R., Szabó, J., Walter, T. (2019). An ensemble prediction model for train delays, *Transportation Research Part C: Emerging Technologies*, Vol. 104, 196-209, doi: 10.1016/j.trc.2019.04.026.
[7] Li, Z.-C., Wen, C., Hu, R., Xu, C., Huang, P., Jiang, X. (2020). Near-term train delay prediction in the Dutch railways network, *International Journal of Rail Transportation*, Vol. 9, No. 6, 520-539, doi: 10.1080/23248378.2020.1843194.
[8] Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., Peng, Q., Xu, X. (2020). Modeling train operation as sequences: A study of delay prediction with operation and weather data, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 141, Article No. 102022, doi: 10.1016/j.tre.2020.102022.
[9] Gao, B., Ou, D., Dong, D., Wu, Y. (2020). A data-driven two-stage prediction model for train primary-delay recovery time, *International Journal of Software Engineering & Knowledge Engineering*, Vol. 30, No. 7, 921-940, doi: 10.1142/S0218194020400124.
[10] Tang, Y., Xu, C., Wen, C., Li, Z., Song, S. (2019). Support vector regression models for delay time predicting considering high-speed rail facility failure, *China Safety Science Journal*, Vol. 29, No. S2, 18-23, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.003.
[11] Zhang, Q., Chen, F., Zhang, T., Yuan, Z.M. (2019). Intelligent prediction and characteristic recognition for joint delay of high speed railway trains, *Acta Automatica Sinica*, Vol. 45, No. 12, 2251-2259, doi: 10.16383/j.aas.c190188.
[12] Hu, R., Xu, C., Feng, Y., Wen, C., Wang, Q. (2019). Prediction of different types of train delay of Guangzhou-Shenzhen high-speed railway, *China Safety Science Journal*, Vol. 29, No. S2, 181-186, doi: 10.16265/j.cnki.issn1003-3033.2019.S2.030.
[13] Zeng, Y., Chen, F., Jin, B. (2019). A prediction model for timetable delays in dispatching area using neural network, *Railway Standard Design*, Vol. 63, No. 3, 148-153, doi: 10.13238/j.issn.1004-2954.201812160002.
[14] Milinković, S., Marković, M., Vesković, S., Ivić, M., Pavlović, N. (2013). A fuzzy Petri net model to estimate train delays, *Simulation Modelling Practice and Theory*, Vol. 33, 144-157, doi: 10.1016/j.simpat.2012.12.005.
[15] Lessan, J., Fu, L., Wen, C. (2019). A hybrid Bayesian network model for predicting delays in train operations, *Computers & Industrial Engineering*, Vol. 127, 1214-1222, doi: 10.1016/j.cie.2018.03.017.
[16] Pullagura, L., Katiravan, J. (2019). Train delay prediction using machine learning, *International Journal of Engineering and Advanced Technology (IJEAT),* Vol. 9, No. 2, 1312-1315, doi: 10.35940/ijeat.A2088.129219.
[17] Huang, P., Wen, C., Fu, L., Peng, Q., Tang, Y. (2020). A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems, *Information Sciences*, Vol. 516, 234-253, doi: 10.1016/j.ins.2019.12.053.
[18] Hansen, I.A., Goverde, R.M.P., van der Meer, D.J. (2010). Online train delay recognition and running time prediction, In: *Proceedings of 13th International IEEE Conference on Intelligent Transportation Systems,* Funchal, Portugal, 1783-1788, doi: 10.1109/ITSC.2010.5625081.