Transportation Research Part C 68 (2016) 350–368

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Modeling railway disruption lengths with Copula Bayesian Networks

Aurelius A. Zilko [a,*], Dorota Kurowicka [a], Rob M.P. Goverde [b]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands
[b] Department of Transport and Planning, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 22 April 2016
Accepted 25 April 2016

Keywords:
Railway disruption
Prediction
Dependence model

ABSTRACT

Decreasing the uncertainty in the lengths of railway disruptions is a major help to disruption management. To assist the Dutch Operational Control Center Rail (OCCR) during disruptions, we propose the Copula Bayesian Network method to construct a disruption length prediction model. Computational efficiency and fast inference features make the method attractive for the OCCR's real-time decision making environment. The method considers the factors influencing the length of a disruption and models the dependence between them to produce a prediction. As an illustration, a model for track circuit (TC) disruptions in the Dutch railway network is presented in this paper. Factors influencing the TC disruption length are considered and a disruption length model is constructed. We show that the resulting model's prediction power is sound and discuss its real-life use and challenges to be tackled in practice.

1. Introduction

It is unavoidable that the operation of a railway system encounters unexpected incidents which disturb or disrupt the railway timetable. Depending on the length of the incident, different measures need to be taken to handle the situation during the down-time. Shorter incidents, usually referred to as *disturbances*, may require only timetable adjustment while longer incidents, usually referred to as *disruptions*, may additionally require rolling stock and crew adjustment. In this paper, we do not distinguish the difference between disturbance and disruption so the term *disruption* is used when referring to these unexpected incidents, regardless of the length.

The topic of disruption management has been a growing research area in the railway operations research. A vast number of different algorithms and models have been proposed for the recovery from the disrupted situation. Cacchiani et al. (2014) provides an overview of these proposed mathematical algorithms and models where some of the presented works mention the uncertainty nature of the disruption length. Given the information about disruption length, the algorithms search an optimal solution to recover from the disrupted situation in the form of timetable rescheduling, rolling stock rescheduling, or crew rescheduling.

However, in reality, disruption length is very uncertain and it is difficult to tell exactly how long a disruption will last. In the Netherlands, this uncertainty creates a big problem that the Operational Control Center Rail (OCCR) in Utrecht faces when a disruption occurs. During this period, train traffic is hindered and timetable is not followed anymore. This is also illustrated as a bathtub model: the normal traffic intensity goes down due to a disruption, stays at a lower level following

* Corresponding author.
E-mail address: A.A.Zilko@tudelft.nl (A.A. Zilko).

http://dx.doi.org/10.1016/j.trc.2016.04.018
0968-090X/

some contingency plan, and recovers to the normal situation after the disruption has ended (Ghaemi and Goverde, 2015). The OCCR's main duty is to handle the situation during this period with the goal of recovering the traffic to normal as soon as possible. The disruption length is crucial piece of information that the OCCR needs. With this information, they make decisions regarding the actions they need to take in order to achieve the goal.

In practice, a series of updated disruption length predictions are made as more information about the disruption is gathered. When a disruption occurs, firstly a rough prediction based on history is made. This is called the "P1" prediction and is taken to be the average of a given disruption length in the past. In the meantime, the mechanics are informed about the disruption and are tasked to repair the problem. From now on, the OCCR is in close communication with the mechanics. After arriving at the site, the mechanics have 15 min to diagnose the problem after which they are required to make a prediction, based on their own judgment, regarding the repair time. This mechanics' prediction is called the "P2" prediction. The mechanics are allowed to update the prediction later on and this updated prediction is called the "P2a" prediction. Finally, when a final prediction can be made, the mechanics are required to update the OCCR with the so-called "P3" prediction. Upon completion of the work, the mechanics inform the OCCR that the problem is solved and the disrupted train traffic at the troubled section can be resumed. Then, they are required to record the information about the disruption with an administrative form that will be stored in a data base called the SAP data base. Unfortunately, this procedure does not include recording the mechanics' own predictions. Moreover, information about the cause of failures is also not required to be in a structured format in the database. The data that is used in this paper comes from this SAP data base.

Thus in current practice, the uncertainty in disruption length is handled by means of a series of predictions based on the mechanics' expertise and judgment. One other way to tackle the uncertain disruption length problem is by representing the disruption length with a probability distribution. Having such distribution allows us to generate random samples of disruption length. This approach is relatively new in railway operation but has been used in several earlier studies in highway traffic engineering. For instance, Golob et al. (1986), Giuliano (1989) and Sullivan (1997) use the lognormal distribution and Nam and Mannering (2000) uses the Weibull distribution. In the railway operation field, Meng and Zhou (2011) models the disruption length in a single track rail line in China with the Normal distribution, while Schranil and Weidmann (2013) models the railway disruption length in Switzerland with the exponential distribution.

In this paper, disruption length is the center of attention. The uncertainty of disruption length is going to be modeled with a probability distribution from historical data. Moreover, several influencing factors of disruption length are considered. The goal is to construct a dependence model between the disruption length and these influencing factors. When a disruption occurs, the model is conditionalized on the realization of the influencing factors resulting in a conditionalized disruption length distribution. This conditionalized distribution represents the disruption length specialized to a specific situation. Then, a disruption length prediction is made from this distribution.

To do this, a proposal made by Zilko et al. (2014) is followed where the dependence model is constructed using the Copula Bayesian Network. The Bayesian Network (BN) technique has been used in transportation research field for several different studies. For instance, Gregoriades and Mouskos (2013) quantify accident risk in road traffic in Cyprus and Chen et al. (2015) constructs a dependence model of travelers' preference for toll road utilization in Texas with BNs. In the railway field, Oukhellou et al. (2008) use the technique to perform broken rail diagnosis.

BN modeling consists of two parts: the graphical structure that represents (conditional) independence in the model and the conditional probabilities between the variables to specify the rest of the relationships. There are different types of BNs depending on how the conditional probabilities are modeled. When all the variables are discrete, as in the case of the three studies mentioned above, the conditional probabilities are modeled using conditional probability tables (CPT). In the case of continuous variables in Gregoriades and Mouskos (2013), the variables are discretized to obtain a fully discrete model. When all the variables are continuous and Normally distributed, the Normal Bayesian Network (which uses the multivariate Normal distribution) can be used.

In this paper, the conditional probabilities between the variables are represented with copula (more on copula in Section 4). Copula is a very useful model for the dependence between continuous variables as this allows the separation of marginal distributions and the dependence. We are interested in the use of copula because our variable of interest, disruption length, is continuous. The use of copula is not completely foreign in transportation research. Srinivas et al. (2006) use several different copula families to model the dependence between vehicle axle weights. Wan and Kornhauser (1997) construct a copula-based model to predict the travel time which is used in a routing decision making problem. Ng and Lo (2013) model the air quality conformity in a transportation networks with copulas.

The disruption length model that we build will assist the OCCR by updating the uncertainty of disruption length every time new information about the disruption is available, in a similar manner with how the "P1", "P2", and "P3" predictions help the OCCR. Moreover, because the output of the model is a probability distribution function, this also gives the OCCR full control regarding which value they want to take as a prediction. Do they want to be more conservative by choosing a value in the upper quantile of the distribution? Or do they want to be more optimistic by choosing a value in the lower quantile of the distribution?

A Copula Bayesian Network model needs to be constructed for each disruption type. To help the model construction, a user-friendly and computationally-efficient software called `UNINET` which implements the algorithm of the Copula Bayesian Network will be used. This software was developed at Delft University of Technology and is available at www.lighttwist.net/wp/uninet. As an illustration in this paper, we construct a disruption length model for disruptions caused by train detection problems in the Netherlands, specifically the GRS track circuit (TC) failures.