DYNAMIC AND REAL-TIME PREDICTION OF DURATION OF INCIDENT

M.Chandesris
SNCF Research Department, Paris, France

ABSTRACT

Unexpected incidents may occur during train circulation and disturb the traffic. When such an incident occurs and cause delays, traffic managers have to be very reactive in order to take the most appropriate decision to reduce perturbation and delays. Presently, French traffic managers share in real-time the information about the events which caused the perturbation and how they are treated.

The aim of this research was to elaborate a real-time prediction for the duration of the incidents. Statistical information on the expected duration of the incident is made available to traffic managers, which will contribute to increase the relevance of their choices in operation recovery decisions.

First, the issues of traffic management is described. Then, incidents and their characteristics are detailed. Two types of models which can be used in such a context are explained in the third part. A part is then dedicated to the elaboration and validation of prediction model. Finally, the real-time prediction of duration of incidents is provided and future prospect are mentioned as a conclusion.

TRAFFIC MANAGEMENT AND DURATION OF INCIDENT

Railway system encounters incidents during operation. An incident is defined as an event which cause or could cause directly or indirectly an interruption or a diminution of the traffic. Numerous causes exist which can delay trains: obstacle on the track, fire around the tracks, disruption of the train, staff delay, etc.

The traffic managers are facing up everyday many incidents, trains have to be re-scheduled in order to optimize the traffic and reduce delays. In each case, the traffic manager has the choice between different solutions, for example he can delay all trains and wait for the end of the incident, but he can also decide to re-route some trains on another track. Such decisions have a major impact on delay propagation, appropriate decision has to be taken in a very short time. In case of dense traffic, each minute counts and delays can grow up very fast.

The French traffic managers work with the EXCALIBUR [2] system which allows to feed incident databases and distribute in real time the information about the management of incident through the SNCF intranet. Maps, description of incidents and actions for their resolution is shared by all people who need it (see Figure 1). The system is also used to consult past incidents and how they have been treated and how long it has taken, how many trains have been delayed, etc.
During operation and real-time operation management, the present system also provides some indication about the duration of the incident by giving the average of the duration times of the past events having the same type. It gives a first information to the traffic manager. Such an information is quite helpful in order to take appropriate decision in re-routing or re-scheduling trains. Indeed, when a traffic manager have to choose to re-route a train or not, his choice depends on the prediction of the duration of the incident. To make it simple : if the incident will be short he retains trains stopped and waits for the soon departure, if not, he re-routes trains. But the information given by the average is not very accurate and does not evolve during the resolution of the incident. The aim of this work is to take profit of large historical databases about incidents in order to improve prediction for the duration of incident and to furnish to the operators dynamic predictive information about the incidents they are managing.

INCIDENT DATA

The system EXCALIBUR provide to operators a large electronic notebook about the life of the incidents. The traffic managers use it in real-time to register incidents they are treating, by doing that, they contribute to filling up a large database of all past events and theirs treatments. The historical database constitute a huge amount of information about incidents : when, where they occurs, how they are treated and their duration time and their consequences on the traffic. In the historical databases, past incidents are described by three types of variables:

1. **The static characteristics** describe the incident. These characteristics are known very soon and are the first one being transmitted to the traffic managers:
   - type of the incident (obstacle on the tracks, waiting for staff, accident with a person, and so on.) and additional information (injured persons, weather condition, fire)
   - localization of the incident (track, train) and characteristics (type of line, department)
   - date of the beginning of the incident (month, day, hour)

2. **The dynamic characteristics** of the incident. These information are collected during the operations and are helpful to follow and evaluate the progressive resolution of the incident :
   - consequence on the traffic (number of blocked tracks, speed limitation)
   - operation in process (type of operation, staff operating)
   - number of gathered pieces of information (indicator of quality of attention)
   - sources of theses pieces of information (driver, station, rescue staff, local traffic managers)
3. **The consequences information.** This third type of information is only available when the incident is "closed", that is when the traffic comes up to "normal":
- total perturbation (number of impacted trains, delays)
- total duration of the incident

Both the static and dynamic characteristics of incidents are known during operation even if some are known progressively. The third type of information is only known when the incident is closed, but it is the key information.

In the present work, a decision support system is developed by furnishing an automatic real-time prediction for the duration of the incident.

**STATIC AND DYNAMIC MODEL**

In order to provide help to operators, we first have to construct a model in order to predict the duration of the incident. A model can be showed as a black-box (cf. Figure 2):

- the input is what we know about the incident at the considered time
- the output is what we want to know: the time to the end of the incident

![Figure 2: Prediction model as a black-box](image)

An another equivalent way to write such a model is to write it as a function F as showed in the equation (1):

\[
\text{Time to the end of the incident} = F(\text{what we know about the incident})
\]

(1)

This model will be fed by static and dynamical information about the incident and will provide a prediction for the duration of the incident. Two types of models were built, depending on the input of the model: static and dynamic models.

**A static model** is feed only by static information and provide an unique estimation of the duration of the incident.

The mean duration of past incidents of the same type is one of the simplest static model. The input of the box is the durations of all past incidents of the same type, equation (1) is now:

\[
\text{Time to the end of incident} = F(\text{durations of all past incident of the same type})
\]

(2)

In our approach, all static variables were included in the model to improve the quality of the prediction:

\[
\text{Time to the end of incident} = F(\text{type of the incident, condition information, localization of the incident, date of the incident})
\]

(3)

**A dynamic model** is feed by both static and dynamical characteristics of the incident. Equation (1) becomes:
Time to the end of incident

\[ T = F(\text{type of the incident}, \text{condition information}, \text{localization of the incident}, \text{date of the incident}, \text{consequence on the traffic}, \text{operation in process}, \text{number and sources of gathered pieces of information}) \]  

Each time a new dynamical information is available, a new prediction for the duration of the incident is computed: it is a real-time prediction.

Dynamic models are more difficult to elaborate because \( F \) has to be more complicated. But, these models are also more useful for the traffic managers because each time a new information about the resolution of the incident is available, a new prediction is provided.

**LEARNING AND VALIDATION PHASE FOR THE PREDICTION MODEL**

Before providing a real-time prediction to traffic management operators, i.e. operating the black-box of the Figure 2, we have to elaborate it. The elaboration of the model is a complex statistical process divided into two main steps. This process deeply relies on the intensive use of the historical database of incidents. For both static and dynamic model, the whole process is the same and it is illustrated in the Figure 3.

![Figure 3: Learn and validate the prediction model](image)

The French national historical incidents database of last years contains all past incidents and their whole characteristics (static, dynamic and consequences).

The first point is to arbitrary divide the historical database into two parts: a learning sample and a validation sample.

During the learning process, we use the whole first part of historical data to build the model. To build the model is to choose the best \( F \) of equation (3) or (4) depending on the model type. The objective is obviously to get the best predictions.
During the validation process, we feed the model built during learning process with the second sample of historical data (only static and dynamical characteristics) to predict duration: we do as if consequences of past incidents are not known. Then, the predictions for the duration of past incidents are compared to true values (known in historical data). If prediction are not good enough, the learning process is modified to obtain a new model, improving prediction quality.

These steps have to be done before the use of the model in operational conditions.

**SOME RESULTS**

Two types of statistical models have been tested: generalized linear model and regression tree model [1].

The **generalized linear model** used to predict the duration is a model with F following an equation of type (5):

\[
\text{Expected duration} = \exp[cst + a_1 1_{\text{incident type=type1}} + b_1 1_{\text{localization of incident=L1+…}}] \tag{5}
\]

In that equation:
- \( 1_{\text{incident type=type1}} = 1 \) if the type of the incident is type1, 0 elsewhere.
- \( \text{cst}, a, b, … \) are the coefficients of the models which have to be estimated during the learning process.

A **regression tree model** may be showed as a series of binary questions:
- does the incident occurs during night?
- if it does, is it a ?
- and so on…

These models can easily be drown as tree, see figure 4.

![Figure 4: Example of (very simplified) regression tree model](image)

The figure 5 shows the error of three static models:
- the median model (“Médiane”): the expected duration is calculated as the median of duration all past incidents with the same type\(^1\),
- the GLM model,
- the regression tree model (“Arbre”).

This figure is read as follows: 80% of predicted duration have an error of prediction (absolute difference between true and predicted values) under \( m \) minutes.

\(^1\) The median is used in order to improve the robustness of the average model.
The new models clearly improve the prediction: the both curves of GLM and regression tree models are over the curve of the median model.

**REAL-TIME PREDICTION OF DURATION OF INCIDENTS**

Once the model has gone through the learning and the validation phase, it can be used under operational conditions: it is directly plugged into the electronic notebook EXCALIBUR. Then, a real-time prediction of the duration of the incident is provided: every time a new incident is created in the EXCALIBUR system, information added, a new prediction for the time to the end of the incident is automatically computed and provided. As seen before, this prediction is acute information for traffic management.

**CONCLUSION AND PROSPECT**

The possibility to provide useful real-time assistance to the traffic managers through a prediction of operation incident characteristics (duration, ...) in order to help the recovery decision was shown. Statistical based models have been built using the whole French national historical incidents database. Predictions provide a useful help to traffic managers and are very encouraging to be improved by adding some more efficient static and dynamic characteristics. Improving predictions of duration of incidents is currently going on by both adding more characteristics and testing new families of statistical models.

This research shows how exploitation of statistical analysis of the historical databases of incidents can be efficient, even in real-time context. This kind of prediction is very useful for the actual and future railway operation system [2] and can help in many different ways:
- to directly provide help to traffic managers
- to feed traffic management tools (automatic tests for re-routing for example)
- to share predictive information for both railway operators and customers.

**ACKNOWLEDGEMENTS**

The author acknowledges A. Lhote and E. Peter for their active participation to the project.
REFERENCES