Modeling random events using a mixed approach Machine Learning techniques

By : Simon Isaac KABEYA MWEPU (PhD student at the University of Kinshasa/DRC)
From : LUBUMBASHI HIGHER INSTITUTE OF STATISTIQUE

Introduction

Traffic accidents are real facts. They are responsible for the death of many people in the world. As zero risk does not exist in transport, the men carry out regular interviews in order to be able to reduce the number of accidents. Maintenance as a method of reducing the number of accidents gives good results in the West, but not in Africa. Hence, we believe that the use of accident prediction would allow us to reduce the number of known accidents ever so slightly. Several studies, focusing more on accidents in road transport, have been carried out around the world using prediction by Machine Learning methods. Wanting to verify the feasibility of such predictions in rail transport, this work consists in predicting the "train derailment" phenomenon using Machine Learning techniques. Thus, knowing in advance, the derailment can be avoided by correcting up stream, the causes mentioned and by evaluating the effects of the corrections made, until ensuring that the phenomenon cannot occur. We believe that such a study will help African railways which in most cases have obsolete equipment. This often results in multi-cause derailments, resulting in multiple casualties. this multiplicity of causes made us think of using Data Lakes as database management tools and the competitive combination of hybrid Neuro-fuzzy classifiers to set up such a model

Methods

The Data Lakes to be set up will contain:

1°) photos of the condition of the track which will allow us to determine the existence or not of a defect in the track likely to cause derailments or not. these photos, taken from different angles will allow us to grasp for each defect: the existence, the quality and the size.

For example :

a)

These two photos illustrate a large joint seen from two different angles (oblique at 60° and 30°

b)

These two photos show show subsidence from different angles. Based on the photos taken on the track, we need to be able to highlight non-apparent defects (e.g.. Under-track voids), defects resulting from a calculation process (e.g.. The warp), whose visibility is not apparent on the surface photos. These photos will be used to qualify any defects that could cause derailments

2°) Maintenance data which are measurements taken from the wagons during maintenance. These measurements are fuzzy (trapezoidal or triangular fuzzy numbers)

3°) Staff data : which can be conductor rating (e.g. bad : [45-50[, poor : [43-57[, fairly good: [55-67[, good:[65-81[, very good:[80-89[, excellent:[88-98[.

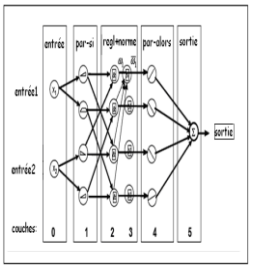
These cotations result from the number of derailments known to the driver. These data are gaussian fuzzy data

4°) Sabotage data which shows the effect of humans sabotaging the track in the occurrence of a derailment e.g. the presence of stones in the interlocking of switches, etc.

The final output is :

- Derailment due to multiple causes : track (X%), equipement (Y%), personnel (Z%), or other (K%), with $X+Y+Z+K+\ldots \leq 100$ %.

- Single-cause derailment : derailment due to track (overgauge, large joint, left on track, warp, etc.), derailment due to material causes (bad brake shoe clearance, brake rod broken and fallen on track, etc.), etc.The data that make up our Data Lakes till be submitted to hybrid Neuro-Fuzzy classifiers such as ANFIS or NEFCLASS (placed in competition) after fuzzification and defuzzification in order to predict the occurrence of a derailment (output)



ANFIS Model

NEFCLASS Model

ANFIS is an adaptative system Neuro-Fuzzy inference system using a five layer MLP type Neural Network from a Tagaki Sugeno fuzzy inference system

NEFCLASS use 3 layers (input, hidden and output) and avoid defuzzification.

These Neuro-fuzzy classifiers will be placed in sequence or in paralle with the possibility of adding others to improve output results.

Results

After applying all these methods, we will obtain the following results :

- before the trains departure, data on maintenance, track faults, drive qualifications, etc. are entered into the classifier, giving the following output ; vehicle n° X lined up for departure will derail for such reason at the given kilometer (K), or vehicle n° Y lined up for departure will not derail.
- After a given period of time, the photos taken on a given stretch of track will provide information on the existence of the « D » defect, at kilometer « Z », enabling the vigilance teams to carry out the appropriate corrective work.

Conclusions

After implementing a model based on the combinaition of Neuro-fuzzy Networks, we hope to significantly reduce incidents in the system, if not eliminate them