A Machine Learning Approach to Automatically Classify Signals on the GSM-R Italian Railways Communication Infrastructure

M. Di Capua(1), **A. Ferraro** (1), **G. Paolocci** (1), **R. Malangone** (2), **L. Di Dio La Leggia** (2).
(1) US srl, (2) Rete Ferroviaria Italiana – RFI SpA
m.dicapua@usmail.it, r.malangone@rfi.it

Abstract

With the increase of the High Speed railways sections and the use of enhanced on-board signaling systems, the current GSM-R infrastructure (Global System for Mobile Communications, used on Railways) for radio communications, between on-board and ground, may not be correctly configured for high service speeds, effectively generating an increase in the number of possible problems as Radio Link Failure (RLF). Consequently, if the performance of the radio communication seriously degrades, this can generate important disruptions (loss of connections with relative loss of control signals) slowing down traffic trains. In order to improve the overall communication system performance, it could be useful to adopt automatic techniques to identify potential critical sections of the railways structure and thus provide useful information for the decision-making process to properly configure the GSM-R network. In this paper we present a currently adopted tool in RFI, named GSM-Resolve, that integrates machine learning methods, based on artificial neural networks (ANNs), to estimate the quality of communication signals, in order to automatically classify and recognize the signals that show anomalies from those considered regular.

1 Background

The Italian and European railway networks, on high-speed and high-capacity lines, have installed the new ERTMS / ETCS signaling system (European Rail Traffic Management System / European Train Control System) born from the need to unify the various national systems signaling, in order to ensure the interoperability of European rail transport. The ERTMS standard defines a fully automated railway traffic control system supported by computerized solutions that communicate and cooperate with each other through the exchange of messages, minimizing human intervention and significantly contributing in increasing the performance of railway traffic in terms of safety, speed, ability and punctuality [Dudoyer *et al.*, 2012]. The current GSM-R infrastructure [He *et al.*, 2016], that is a subsystem of ERTMS, is used for radio communications between trains and railways and control centers, may not be correctly configured for high service speeds, effectively generating an increase in the number of unwanted handovers events (i.e. Ping Pong effect and Radio Link Failure (RLF)). In order to monitor these communications links the company US srl, together with RFI SpA, has developed the tool GSM-Resolve, a diagnostic software, currently in use, which provides support for the identification of possible anomalies on the railway network and which is of fundamental help in resolving the disputes between the parties involved (network managers or user companies). More specifically, the system allows you to analyze the radio messages that pass through the GSM-R infrastructure in order to detect any unexpected behavior that is not in line with the operating logic defined by the protocols used by the communication network. The tool analyzes in particular the ABIS communication interface (BTS-BSC), the A interface (BSC-MSC) and the ISDN interface (MSC- RBC). Currently the software provides a variety of information and indicators based on rules defined by TLC and railway signaling experts, which are then applied to the input data and shown through dashboards and graphs, which however must be interpreted by experts in order to establish the possible fault root cause [Zhou et al., 2014]. The innovation recently introduced on the GSM-Resolve system concerns the development of an additional diagnostic component, based on machine learning techniques, which is able to automatically classify and recognize the communication signals that show anomalies from those considered regular.

1.1 Communications over GSM network problems

In the most common configuration of the ERTMS system, the exchange of information between the “ground” and on-board subsystem takes place thanks to the GSM-R radio channel (where R stands for Railways), which allows voice, SMS and data communications between the personnel of the Infrastructure Manager and Railway Companies, providing specific functions such as functional number call, group calls, and railway emergency calls. In addition, in railway lines equipped with ERTMS-based signaling and traffic control, the GSM network provides the transmission data channel for communication between the on-board and ground-based Command, Control and Signaling technological systems. Actually, the considerable increase in the speed of modern trains today causes frequent *handhover* events between cells, i.e. changes in the channel used by the connection of a mobile terminal to the GSM network while maintaining communication active. Given the size of some cells of 1-2 km, a high-speed train traveling at 350 km/h generates handover events with a frequency of 10-20 seconds. Solving the problem of effectively controlling the frequent rides (messages) between cells is one of the main functions of current railways management systems whose diagnostics represent the future challenge of high-speed rail systems.

It is evident that to ensure a reliable rail service, radio equipment must be kept in good working order through regular checks. The long-lasting and efficient operation of railways depends on the ability to identify any faults in the network equipment at an early stage. Events such as stopping a train can have immediate implications, for example the blocking of rail transport, with related timetable changes and unmanageable consequences. The management and eventual prevention of such failures is currently delegated to diagnostic systems mostly made privately, which tend to analyze the failure scenarios in a timely manner, mainly interpreting the possible causes through "human" intervention [Xu *et al*., 2016]. The intrinsic limitation of current diagnostic systems mainly concerns the lack of approaches aimed at using the large amount of data produced by control systems with advanced techniques for analyzing such data. Despite the fact that the adoption of advanced analysis techniques, such as machine learning, can improve the performance of a railway transport system, there are still few companies linked to the railway sector that have implemented these techniques in one or more transport sectors. railway. This is mainly due to the lack of understanding of the techniques to be implemented and the inability to properly collect, process and structure the mass data produced by the systems [Moreno *et al.*, 2015].

2 The GSM-R tool

US s.r.l. has already developed and tested on the field, together with RFI (Rete Ferroviaria Italiana) the GSM-Resolve software, a monitoring and diagnostic system that analyzes the radio messages passed through the GSM-R infrastructure, in order to detect any unexpected behavior not in line with the operating logic defined by the protocols used by the communication network. In particular, the tool analyzes the ABIS interface (BTS-BSC), the A interface (BSC-MSC) and the ISDN interface (MSC-RBC). For each anomaly detected, a report accompanied by information on the possible causes is sent to a central control system in order to make it available to any other connected subsystems. The tool integrates an open database of rules, defined by TLC and railway signaling experts, that are applied to the data flows received by the tool as input. The GSM-Resolve system allows experts to gain an overview of each ride, reporting all the GSM calls and any anomalies associated with a specific call at a given moment. It is possible to choose whether to display the time parameter or the train position parameter on the abscissa, while all the GSM users involved are shown on the ordinates. By GSM users we mean both the two on-board radios and all the railway personnel connected at that moment to the GSM-R network (train conductor, driver, etc.). Each call is graphically represented by a segment, with exact information about the start and the end of the call, the reason for the termination and all the numerical anomalies found by the system, highlighted with different colors based on the severity. From any anomaly, the system is able to display the entire sequence of messages relating to the call, highlighting the message that triggered the selected anomaly. For each call or for the entire route, it is possible to inspect all the BTSs, in a time sequence and also with some information such as: signal level, time advance, and quality. In the following figure 1 it is shown just a screenshot of the GSM-Resolve system, showing the different GSM signal levels, recorded for each cell crossed by the train in a specific section of the route.

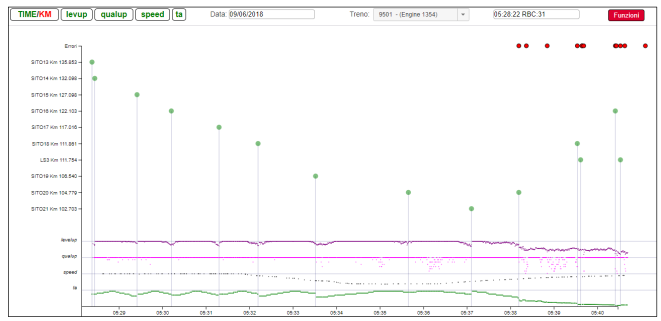

Figure 1. Different GSM signals levels
showed by the GSM-R tool.

3 Methods: An ML approach in GSM-R

The signals managed by the GSM-Resolve system are essentially made up of numerical sequences whose values represent the levels of the recorded signal. The trend of the signal over time (ascent or descent) plus some defined variables (and thresholds) represent input data that can feed an artificial neural network model, suitably trained with sufficient training data, in order to automatically extract relevant features of the signal and classify it. From a methodological perspective, the type of training adopted is the supervised one, that is a training process in which data (sequences of values) already labeled in terms of correct / anomalous signal are passed to the model. The neural network model adopted in GSM-R for the automatic classification of GSM signals is based on the so-called Multi-Layer Perceptron network (MLP), see figure 2. The reasons behind the choice of this model fall mainly on the fact that our goal can be compared to the one of finding an approximation of functions to one or more variables, of curves and surfaces that can be effectively solved with a feed forward neural network and does not require neural networks with complex architectures, in order to obtain robust results: an MLP (a multilayer perceptron) is sufficient where the input layer contains as many neurons as the size of the domain and the output layer contains as many neurons as the size of the codomain while there are some options in choosing the architecture of the hidden layers, the activation functions, the loss functions, the optimizer and the various training parameters.

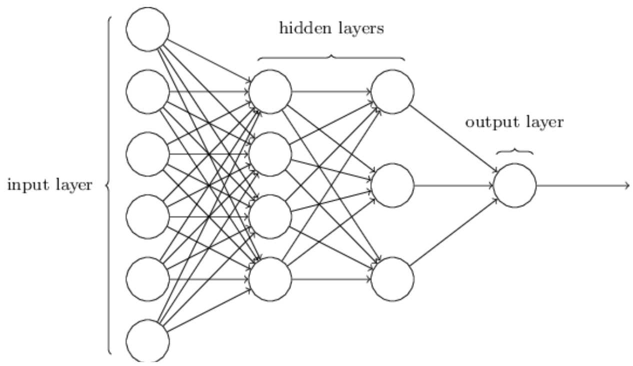

Figure 2. A general MLP architecture schema.

In our case, an initial model was designed with a first input layer made of 100 neurons, connected to a subsequent dense layer (i.e. composed of neurons fully connected between the two layers) with 256 neurons. The inner hidden layers are still dense layers with 256 up to 512 neurons, which finally connects to an output layer which has only one output neuron, being the goal of the solution a binary classification of the input. The output of the network (output neuron) has been managed with the sigmoid function, that is a mathematical function that produces a curve called sigmoid, having an "S" shape: a function commonly used in binary classification tasks. The loss function, which represents the network error obtained during training, and which must be "minimized" by the training algorithm (in our case is based on the classic "backpropagation" algorithm), is represented by the Binary Cross Entropy function. The optimizer adopted in the training phase of the network is instead based on ADAM, an optimization algorithm that is an extension of the stochastic gradient descent, recently widely adopted for learning applications in deep architectures.

4 Results

GSM signals (levels) are mainly numerical values representing the intensity of the radio signal. In the figure 3 above it's possible to see an example of a bad signal, with on the y-axis reported the intensity of the signal over time (x-axis), in 100 subsequent samplings.

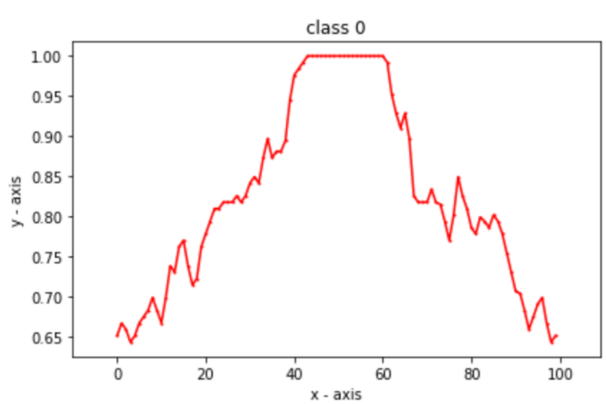

Figure 3. Sample of an anomalous quality signal in GSM comm.

The signal in figure 3 shows a "gaussian shape", denoting a good quality only in the central part, not sufficient for a good GSM call. In figure 4 is showed another bad signal, that starts with an evident low intensity (above 0.65) and increases its quality, only for a certain period, ending with a decreasing intensity.

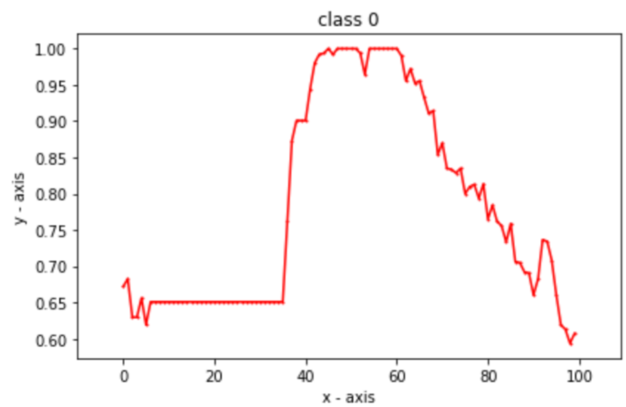

Figure 4. Another sample of an bad quality signal in GSM comm.

A sample of a good signal is instead reported in figure 5, where it is possible to see an overall high intensity of the signal, with only a loss in the last part of the sequence.

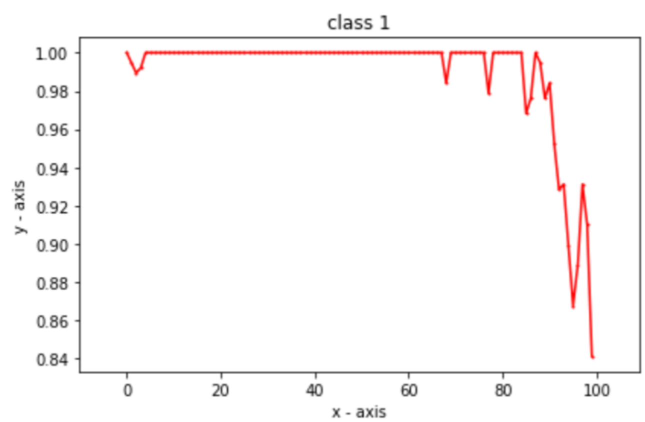

Figure 5. Sample of a good quality signal in GSM comm.

4.1 Experimental setup

We collected 3 different bags of signals, registered in different geographical routes. Each dataset contains about 50k signals, balanced between normal (class 1) and anomalous signal (class 0), labeled in a semi-automatic manner with the help of GSM communication experts. No data preprocessing has been applied in the experimental setup, except for the normalization of data values, between 0 and 1. The model has been trained on 500 epochs, with a K-fold cross validation (K=10), to avoid possible overfitting problems. The selected loss function is the Binary Cross Entropy, that represents the measure of uncertainty associate with a given distribution *q(y)*. The corresponding equation is illustrated above:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

where *y* is the class label and *p(y)* is the probability predicted by the model.

The frameworks used for the design and train/test activities are Keras and TensorFlow2. At the best of our knowledge, given the specificity of the problem and the context (GSM-R), there are no similar case studies to compare with, while some research are made on the 4G network and above [Savas et al., 2012], so mainly we prepared 3 *in-house* datasets to test the efficacy of the proposed model. In the table 1 the test accuracy metric has been applied and compared among 3 datasets.

Dataset	Loss	Accuracy
A	0.36	0.87
B	0.25	0.88
C	0.30	**0.89**

Table 1: Classification accuracy results.

The average accuracy obtained by our initial model is largely above the 85%. Our preliminary test show that the provided solution is promising in order to support the integration of automated diagnostic tools, as GSM-Resolve, based on machine learning approaches, within a certain efficacy, with the aim not only of being able to classify communication faults but also of being able to predict them.

5 Discussion and further directions

In the current monitoring systems of railway communication equipment, automatic fault classification activities are not really addressed, and the waveform of the GSM signal is analyzed together with other measurable values, to derive hypotheses on possible failures to the related infrastructure part [Moreno Garcia-Loygorri *et al*., 2018]. In order to automatically classify failures in the communication system of the GSM-R network, or to predict failures in the system itself, it is necessary in terms of objectives, to design and implement innovative algorithms capable of automatically learning the correct network behaviors. communication, i.e. all operating scenarios that are in line with what is established by the TLC experts, and therefore be able to discriminate with respect to behaviors that are instead typical of potential malfunctions. More in detail, in line with the current themes of Industry 4.0, in the described solution we adopt machine learning techniques in order to develop innovative diagnostic components. The proposed solution allow specifically the integration of artificial neural networks in diagnostic tools, in order to automatically learn patterns underlying the data, without being explicitly programmed. Even if it was not possible to compare our results with similar approaches in literature, at least for GSM signals in the railways field, the preliminary results are promising and there is space to improve further the accuracy of the classification, also exploring predictive models that can strongly mitigate derived train malfunctions.

In the next future, among the different models that can be tested in this context, we underline the possible adoption of LSTM models (Long Short-Term Memory). LSTM networks are one of the architectures, starting from recurrent neural networks (RNN), that has received more attention in recent years from the scientific community. One of the main advantages of LSTMs is the ability to learn from long time sequences, using memory to avoid lags of unknown duration, between important events in a time series. Applying LSTMs for diagnostic systems can allow to monitor the entire history of the "degradation" process and consequently correctly predict any faults. Apart from a safety perspective, also from a business perspective recent studies estimate how the effective use of predictive maintenance in railways can reduce downtime by up to 50%, with an estimated savings of 10 to 40% on equipment maintenance costs.

References

[Dudoyer *et al*., 2012] *Susceptibility of the GSM-R transmissions to the railway electromagnetic environment*. In Signaling and Security in Railway; Perpinya, X., Ed.; IntechOpen: London, UK, 2012; pp. 503–522.

[He *et al*., 2016] *High-Speed Railway Communications: From GSM-R to LTE-R*. IEEE Veh. Technol. Mag. 2016, 11, 49–58.

[Moreno *et al*., 2015] *A survey on future railway radio communications services: challenges and opportunities*. In *IEEE Communications Magazine*, vol. 53, no. 10, pp. 62-68, October 2015.

[Moreno Garcia-Loygorri *et al*., 2018] *The Wireless Train Communication Network: Roll2Rail Vision*, in *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 135-143, Sept. 2018, doi: 10.1109/MVT.2018.2844408.

[Savas *et al*., 2012] *Analysis of Mobile Communication Signals with Frequency Analysis Method*, Gazi University Journal of Science, vol. 25(2), PP.447-454, 2012.

[Xu *et al*., 2016] *A survey on high-speed railway communications: A radio resource management perspective,* Computer Communications, Volume 86, 2016, Pages 12-28, ISSN 0140-3664.

[Zhou *et al*., 2014] *Provide High-QoS of the High-Speed Railway Mobile Communications in Cyber-Physical Systems*. Computer Science, 3 September 2014. Arxiv.