Firemen prediction by using neural networks: a real case study

Christophe Guyeux[1], Jean-Marc Nicod[1], Christophe Varnier[1], Zeina Al Masry[1], Nourredine Zerhouny[1], Nabil Omri[1], and Guillaume Royer[2]

[1] Femto-ST Institute, UMR 6174 CNRS, University Bourgogne Franche-Comté, France
[2] SDIS25 - Service Départemental d'Incendie et de Secours du Doubs, France

Abstract. Being able to predict the daily activity of firefighters is of great interest to optimize human and material resources. It will allow to enable a quicker response by achieving a better geographical deployment of these resources according to the expected number of interventions. Having obtained the list of interventions for the period 2012-2017 in the Department of the Doubs, France, we added a relevant collection of explanatory variables based on calendar data (time of day, day of the week, day of the month, year, public holidays, etc.), road traffic, meteorological and astronomical data, and so on. After detecting outliers and completing missing data, this set has been divided for learning, validating, and testing. The learning is then carried out on an ad hoc multilayer perceptron whose hyperparameters are finely defined using some supercomputer facilities. This neural architecture are finally applied on a real case study, that is, to the predictions of firemen interventions for the year 2017 after a learning stage on 2012-2016, leading to really encouraging results.

1 Introduction

It is reasonable to assume that firefighters' interventions are not purely random, but are at least partly conditioned by climatic, temporal and other events. Based on this principle, we should be able to find features that make it possible, to a certain extent, to predict how many interventions they will have to carry out in the coming hours, in the coming days, etc. At a time when human and financial resources are constantly being reduced, among others due to the international financial crisis and societal choices, when demand is steadily increasing in France (closure of local hospitals, population growth and ageing, etc.), better control and optimization of the use of its human and material resources is essential. Building a tool to predict the number of interventions can be a great help in this search to optimize the response in constant resource. Indeed, on a constant basis, if it is possible to ensure that the number of firefighters on duty can be sized so that the periods of duty with the most firefighters available correspond to the periods when they are most frequently used. Then, the number of situations of breakdowns with speed objectives, as well as the number of situations where a centre can no longer respond (to a fire or personal assistance) due to lack of resources, will be reduced.

Such a prediction tool seems feasible, and has already been tried punctually [10,11], due to the very nature of firefighting operations. Indeed, the latter is first of all directly related to human activity. For example, there is generally less intervention at 2 a.m. than at 2 p.m., because in the first case the majority of humans, the source of fires and accidents, sleeps. Similarly, in France, the lunch break is generally well followed, leading to a decrease in activity, and therefore to fewer accidents (at work, on the roads, etc.) And there are probably no accidents at work on public holidays, when weekends see an increase in leisure-related interventions: drownings in swimming pools, road accidents linked to excessive alcohol consumption on weekend evenings, etc.

On the one hand, there is therefore a number of interventions per hour that clearly depends, in part, on external events such as time, month, human activity, climatic conditions, etc. On the other hand, fire brigades are used to record response data (begin and end times of each intervention, its locality, etc.) for various legal reasons and for activity statistics. It therefore seems possible and very interesting to collect a set of explanatory variables potentially related to the variable to be explained (in this paper, the number of interventions per hour). These data would allow supervised learning, using proven techniques such as neural networks, to provide a predictive tool for the number of interventions. The objective of this article is precisely to show that such a predictive tool is possible on a concrete case study (the interventions of the Doubs fire brigade, France). We will show how, from the list of interventions over the period 2012-2017, it is possible to build a set of explanatory variables (features) and to set up a neural network, so that automatic learning is possible, and leads to a tool that provides predictions reasonably close to the actual number of interventions.

The remainder of this article is organized as follows. The problem considered in this article is presented in the next section, while the dataset construction and exploitation is detailed in Section 3. Then, in Section 4, a multilayer perceptron designed with Tensorflow is proposed with full details. Its preformance is deeply evaluated in a real case study in Section 5. This research article ends by a conclusion section, in which the contributions are summarized and intended future work is outlined.

2 Problem statement

Since 2012, the Doubs fire brigade has been collecting the date and time of each of their interventions within the fire department. This represents about 200,000 interventions. For each intervention, we only know four features: (i) the reason for leaving, (ii) the date, (iii) the time of departure and (iv) the return to the base station. Our objective in this work is to be able to predict hourly the 2017 interventions (test set), based on a learning and validation over the 2012-2016 period.

The number of interventions does indeed have several seasonal trends, as can be seen in a seasonal decomposition using moving averages (additive model, cf. Figure 1). It is related to human activity, punctuated by our biological clock (less activity at night, because most of us sleep), our cultural heritage (Sunday, holidays), and the alternation of seasons (ice storm in winter, fires in summer). The year, for its part, weighs the number of interventions for societal reasons: the cumulative effects of various unclear reasons (hospital closures, global warming involving more floods and fires, the ageing of the population, etc.) lead to an increasing number of interventions from year to year. These four features (hour, day in the week, month and year) therefore make it possible to recover the skeleton of the seasonal component (alternating hours, days in the week, and month) as well as the general upward trend (via the year).

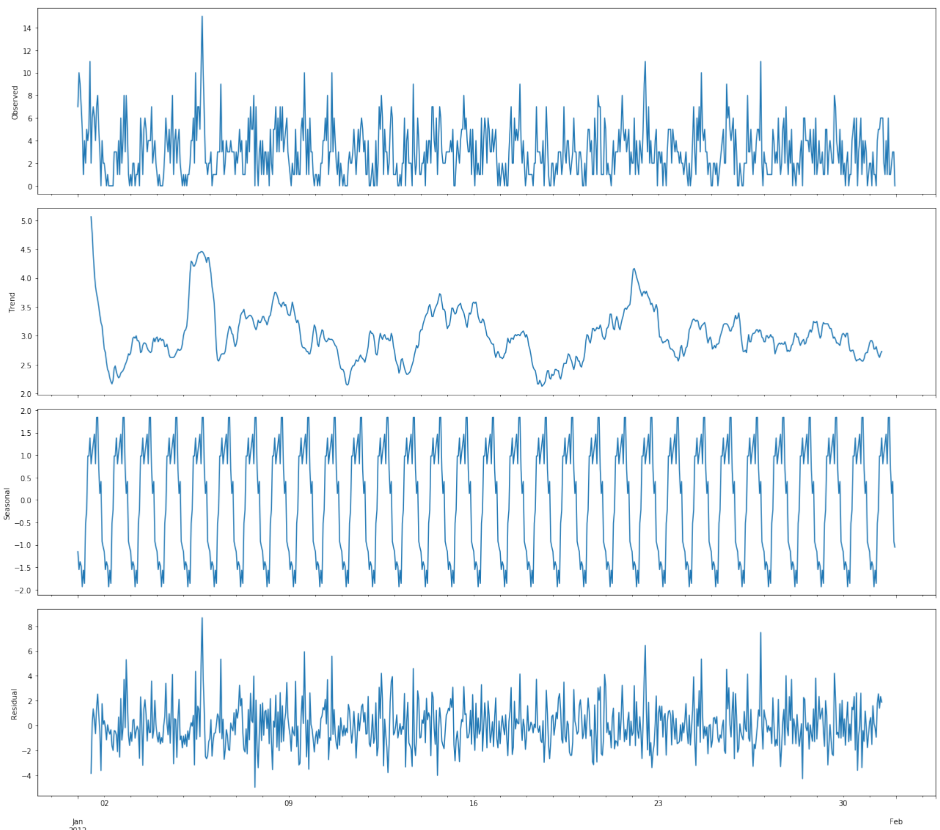


Fig. 1: Firemen interventions: seasonal tendencies

The first task to do is to build a basis of knowledge by increasing the number of features. A list of potential measurable reasons for the variability in the number of interventions should be produced after careful consideration. Data should also be collected from different sources related to these explanatory variables, leading to enhanced features. In doing so, false positive alerts may be introduced, and prospective links may ultimately prove to be uncorrelated to the number of interventions. This leads to an unnecessary complexity of the model (and therefore to increase the learning difficulty), and to additional noises. For example, should the phases of the moon be added, in accordance with a popular belief shared by firefighters, that the number of births increases during full moon nights? The database must then be cleaned of its outliers, completed for missing data, and formatted: normalization for numerical features, and one-hot-encoding for categorical ones.

The next step is to choose a predictive model appropriate to the complexity of the data. While it is true that the activity of firefighters is highly dependent on a certain number of variables, the latter are potentially large in number, some of them only influencing the number of interventions at the margin. For example, there is a peak in suicides on June 21 [14], which tends to slightly increase the number of interventions that day. Similarly, flooding in the rivers can lead to an increase number of people rescued, when during chicken pox outbreaks, ambulance attendants are saturated, and people must therefore call on the fire brigade to provide backup. These relationships to the number of interventions are subtle and complex, and require the introduction of many explanatory variables. As illustrated in Figure 2, traditional machine learning tools such as vector machine supports cannot produce predictions that meet expected results, neither with a small number of features (due to the lack of representativeness of this too small data set), nor with a larger number (due to the complexity of the link). This is why, to be well understood, these data require the use of deeper architectures such as neural networks that know how to undertake them.

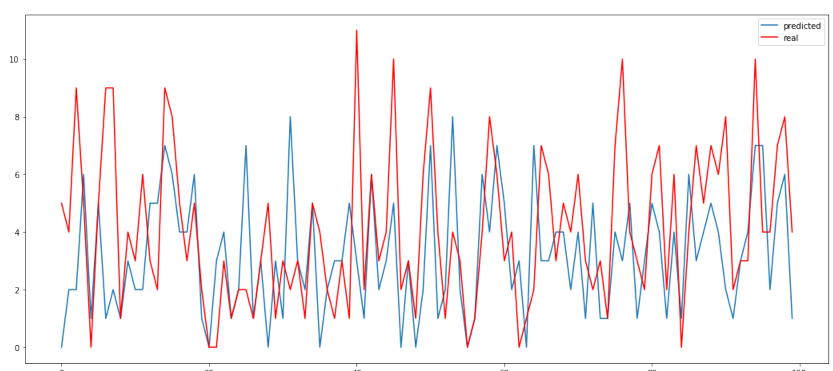


Fig. 2: Bad performance of the SVM predictions, even after hyperparameter fitting: hours randomly picked in test set on the abscissa, and number of interventions in ordinate.

At this stage, the problem becomes to find the best neural architecture (kind of neural network, number of layers, number of neurons per layer, kind of activation function, etc.) and to apply it on the best subset of features. These model selection and hyperparameter fitting is the objectives of Section 4, while the next section focus on feature collection.

3 Dataset construction and exploitation

Apart from the four previous features that have an obvious impact on the general trend and the seasonal part (time, day in the week, month and year), it seems obvious that some days in the year are not neutral, but potentially generate an increase in interventions. Examples include Christmas and New Year's Eve, the French National Day (July 14, with its fireworks and associated risks), or the summer solstice and its increase in suicides. Day in the year therefore is an important explanatory variable, which poses an encoding problem. Indeed, encoding it as a numerical feature through standardization leads to a weak distinction between 23 and 24 December, whereas there is a gap in interventions. Conversely, using categorical one-hot-encoding raises two problems. First of all, due to leap years, 21 June does not always fall on the same day in the year (two possible solutions: either delete all 29 February or add a qualitative variable corresponding to the leap character). Then, this encoding will introduce 365 new binary features, which will pose a problem for basic regression tools. One-hot-encoding was nevertheless chosen, because it makes it possible to highlight certain very particular days for certain types of intervention.

Other special days are school holidays and public ones, which obviously have an impact on human activity. Various effects of school holidays on the number of interventions can be reported. For example, not being at school, children ride more bicycles, swim in swimming pools in good weather, and go for walks in the forest more often (the Doubs being heavily wooded), and all these activities are accident-prone. On the other hand, the Doubs is probably less populated during the holidays, being a relatively low tourist destination: in winter, the inhabitants tend to go on ski holidays in the Alps, when in summer they are more often found at the beach elsewhere. Thus, if the activities are more accident-prone during school holidays, they are inversely less important due to a certain decrease in the population during these periods, and the trade-off between the two trends is difficult to measure.

What seems certain is that the end of the day on wakes and the end of holidays generate more traffic, more agitation, and therefore probably more accidents. To this end, Boolean features have been added corresponding to being on holiday, and being on the eve or on the return of the latter. The values of these features, which are few in number (five periods per year), were entered by hand from the academic calendar found on the website of the French Ministry of Education [4]. As road traffic is likely to be important, the orange, red and black alerts for dense regional traffic have also been integrated, collected from the government website "bison-futé" [3], which led to numerous qualitative variables.

Finally, public holidays are only for some of them related to school holidays, and do not always fall on the same day in the year: if the national holiday and New Year's Day fall each year on the same day, the same cannot be said for mobile religious holidays (Easter, ascension, etc.). Also, these particular days gave rise to a new qualitative variable, which could be easily defined using some Python work calendar module.

Another category of features necessarily has an impact on the number of interventions, namely meteorological data. Indeed, when it snows, car accidents are more frequent, and motorcycle accidents are more frequent when it rains, etc., but people tend to go out less too. In freezing weather, there are pedestrian falls, when fires are more frequent in very hot weather. When the weather is nice, people have more barbecues or use more swimming pools, and when it rains heavily, it leads to floods and then to personal assistance. There are therefore many examples where climatic reasons may well explain firefighters' outings. Another advantage of these features is that, as with holidays and vacations, they can be predicted: 2018 data can be used first to predict the weather in 2019, then refined as the days approach, with updates to the national weather forecast system.

In France, meteorological data archives can be found on the Météo-France website [5]. They are available until January 1996, and for a collection of weather stations distributed throughout France. Data have been collected from Nancy, Dijon, and Mulhouse, the three closest stations to our region. For each station, the following quantities were considered: temperatures, sea level pressure, local pressure, pressure variation in 3 hours, barometric trend type, total cloudiness, humidity, dew point, last hour precipitation, last 3 hours precipitation, average wind speed 10 minutes, average wind direction 10 minutes, gusts over a period, horizontal visibility, and present time (qualitative data). All these data, except the last one, have been standardized. It should be noted that, on the one hand, meteorological data are only available every 3 hours, and that some data were missing for some 3-hour ranges. And so a linear interpolation had to be implemented, after checking that these gaps are not too frequent.

Finally, ephemeris data have been added such as the time of sunrise and sunset (more precisely, the dawn and dusk hours in Besançon, department of the Doubs), as well as the time of moonrise and moonset, with its phase. This has a direct impact on drivers' visibility, when there are few roads lit at night in France. Such features have been added to take into account the fact that, if the moon is full and rising, and there are no clouds (cloudiness feature already integrated), then visibility is improved at night. In a more anecdotal way, we have access to the "full moon" data, which allows us to see if it influences certain types of interventions.

The results of this dataset construction is a matrix of 747 features, for 52,560 rows (one row per hour over 6 years, between 2012 and 2017), leading to 39,262,320 data to deal with.

4 A MLP for firemen: architecture design

The neural network was built using the sequential model from the Keras module, Python TensorFlow backend [1]. To select the basic neuron, the ELU [2], RELU, and SELU (Scaled Exponential Linear Units [9]) activation functions have been tested in various network architectures, and the SELU function has at each time produced the best root mean squared error (RMSE) on the test set. The He normal initializer has consequently been chosen [6], this initialization being well adapted for the SELU activation function. Similarly, two optimizers have been tested on this dataset and on various architectures, namely the stochastic gradient descent (with or without Nesterov, momentum of 0.9) and the Adam optimizer [8], with or without the AMSGrad variant [12]. In view of the results obtained, we finally chose Adam without his variant, while the number of samples per gradient update (batch size) was specified to 50.

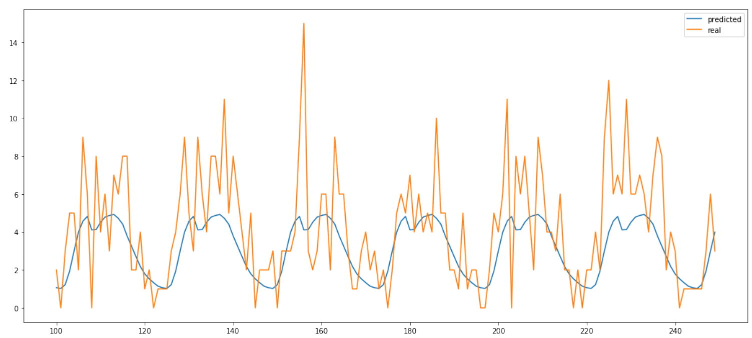
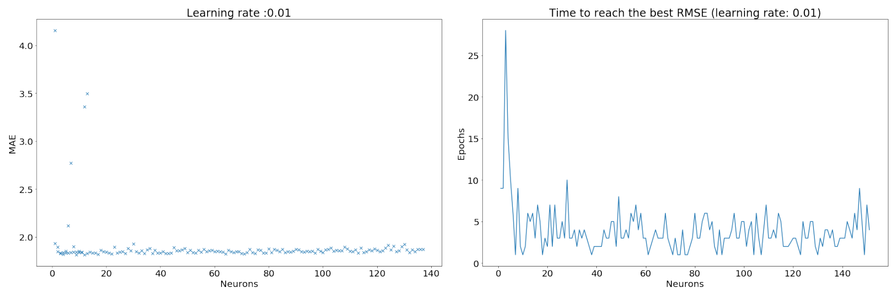
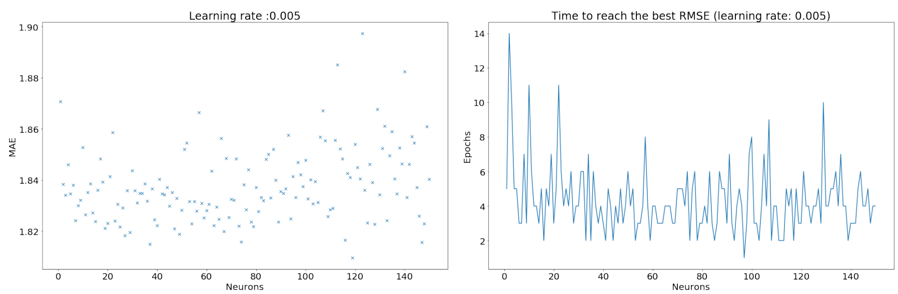


Fig. 3: Real number of interventions versus per-hour averages in the first hours of 2017.

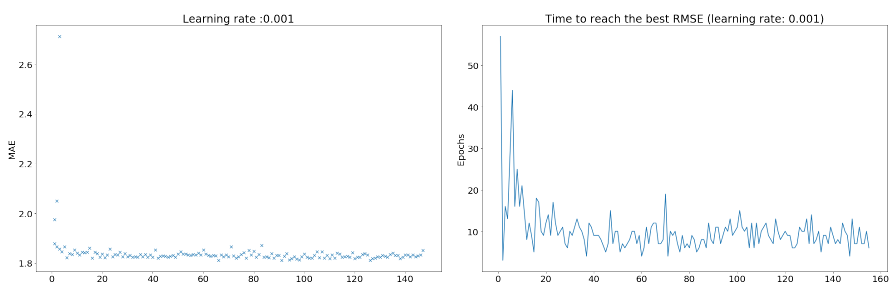
The number of network inputs corresponds to the number of features, and the output corresponds to the number of interventions. We tested one hidden layer, then two. For the first hidden layer, a number of neurons ranging from 1 to the number of features (747) has been tested, and for a learning rate of 0.01, 0.005, 0.001, 0.005, 0.005, 0.0001, 1e-5, and 1e-6. The best first layer was then memorized (made up of 119 neurons, learning rate of 0.005 for 93,177 parameters) with its weights, which we reused as it was within a network of two hidden layers. The first layer was not trainable, while for the second, we tested a number of neurons ranging from 1 to the number of features. Again, the entire list of learning rates above has been tested for this second layer. Finally, Dropout has been considered to prevent this neural network from overfitting [13] and Alpha Dropout (drop probability of 0.1) has been chosen to keep mean and variance of inputs to their original values, in order to ensure the self-normalizing property even after dropout, as advised in [9]: this dropout fits well to SELU by randomly setting activations to the negative saturation value. Finally, two callbacks have been applied to determine the number of epochs during training: an early stopping of training once the loss starts to increase (patience set up to 100 epochs) coupled with a model checkpoint to store the best weights once satisfying the early stopping criterion.



(a) Learning rate = 0.01



(b) Learning rate = 0.005



(c) Learning rate = 0.001

Fig. 4: Time to reach the best RMSE for larger learning rates

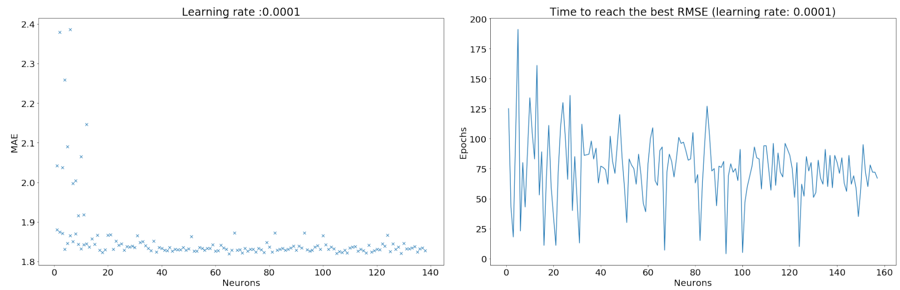
5 Obtained results

First of all, let us introduce some reference values, which will then allow us to evaluate the performance of our neural networks as part of the prediction of the number of interventions. Two reference values are indeed easily accessible, namely to use the average number of hourly interventions over 2012-2017, and the average per hour, as shown in Figure 3. In the first case, *i.e.* by predicting at each hour that there will be 3.3705 interventions, the mean absolute error (MAE) is 2.4278 for a RMSE of 3.1613. In the second case, these errors are reduced to 1.9021 and 3.1613 respectively. An artificial intelligence-based tool for predicting the number of interventions is only useful if it allows to go below these reference errors.

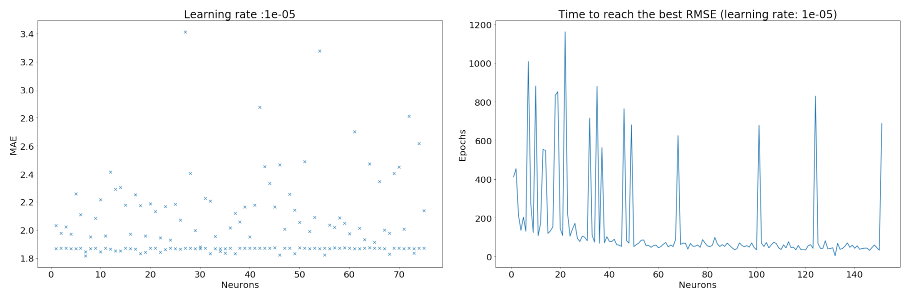
The architecture test protocol mentioned in the previous section was implemented on our supercomputer, which first of all made it possible to dimension the first hidden layer. As can be seen in the Figure 4a, with a learning rate of 0.01, the smallest RMSE is very quickly obtained, in only a few epochs, and the associated MAE of 1.81291 (12 neurons) allows to do better than the average number of interventions per hour. It is possible to improve these scores by enlarging the number of neurons while reducing the learning rate, as can be seen in Figure 4b. The resulting MAEs are more diffuse, while a slightly larger number of epochs are required before overfitting occurs. With such a learning rate, we are able to achieve a MAE of 1.8095 with 119 neurons. These scores remain essentially the same by decreasing the learning rate to 0.001, see Figure 4c (MAE = 1.81005 for 126 neurons), but start to become less good when the learning rate is lowered again. This indicates that, in this situation, we fall into local minima that we can no longer leave. Indeed, the best scores are respectively 1.81886 (152 neurons), 1.81446 (7), and 1.82761 (29) for learning rates of 0.0001, 1e-05, and 1e-06 respectively, see Figure 5.

As stated previously, the best model with one layer has been reached with a learning rate of 0.005 and 119 neurons. Its ability to predict 2017 is depicted in Figure 6. This neural network was then reused, with its stored weights, in another network having a second hidden layer, as indicated above. The neurons added to this second layer were configured in the same way as for the first layer (SELU activation function, normal He initialization), and an alpha dropout of 0.1 was also added. Finally, the number of neurons for this second layer varied from 1 to the number of features. However, in many situations the score obtained was worse than with a single layer and, if finally we managed to do better with a second layer of 54 neurons, the improvement was only marginal (6% at the RMSE level).

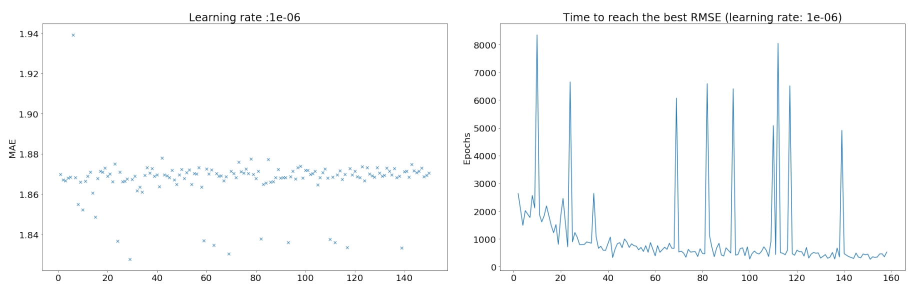
Going into more detail, we realized that 2017 was a very special year, in that private ambulance drivers chose that year to withdraw from a number of low-profit interventions, which were normally their responsibility in previous years. In the absence of a private ambulance, patients therefore refer to the fire brigade. This has resulted in a significant increase in the number of interventions for the year 2017 and, on the one hand, this increase cannot be learned from 2012-2016 onwards and, on the other hand, these taxi operations no longer have much to



(a) Learning rate = 0.1e-4



(b) Learning rate = 1e-5



(c) Learning rate = 1e-6

Fig. 5: Time to reach the best RMSE for smaller learning rates

do with interventions generated by accidents, fires. In particular, they are no longer highly correlated with climate data, etc. If, instead of predicting 2017 from 2012-2016 onwards, another year was predicted, the performance of the neural network presented above would improve significantly, reaching an MAE of 1.2658 by 2012.

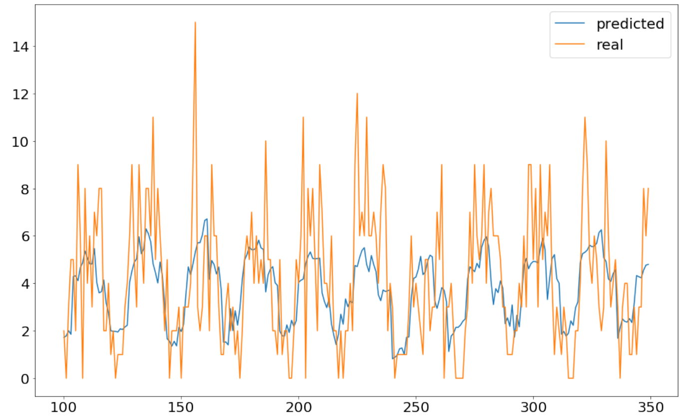


Fig. 6: Predictions versus reality by using the best neural network with one hidden layer.

6 Conclusion

A case study based on real intervention data from the Doubs fire brigade has been studied in this article. From the simple list of interventions over the 2012-2017 period, tools have been put in place to collect 747 features related to meteorology, calendar and road traffic information, etc, and these have been transferred to a neural network for learning purposes. Various ways of building this network were studied, including the choice of the activation function, the optimizer, the dropout, as well as the number of layers and neurons per layer.

The first results obtained by this MLP are really better than our baseline for 2012, but they are not exceptional for 2017. This first brute force approach to determining the right neural architecture, by systematically increasing the number of neurons and then the number of hidden layers, has its limits. Indeed, this gluttonous approach (finding the best first layer, then the best second layer) does not necessarily lead to the overall optimum (the best two-layer), especially since a certain number of couples (learning rate, number of neurons) produced scores close to the optimum. This is why a different approach for the discovery of the best neural architecture for predicting the number of firefighters' interventions should be investigated in a future work, encompassing recurrent architectures like the LSTM ones [7].

References

1. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster,

Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

2. Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *CoRR*, abs/1511.07289, 2015.
3. Ministère de la transition écologique et solidaire. Bison futé. urlhttps://www.bison-fute.gouv.fr/. Accessed: 2018-11-18.
4. Ministère de l'éducation nationale. Calendrier scolaire, france. `http://www.education.gouv.fr/pid25058/le-calendrier-scolaire.html`. Accessed: 2018-11-18.
5. Meteo France. Données publiques. `https://donneespubliques.meteofrance.fr/?fond=product&id_product=90&id_rubrique=32`. Accessed: 2018-11-18.
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.
7. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
8. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
9. Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *CoRR*, abs/1706.02515, 2017.
10. Michael Madaio, Shang-Tse Chen, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng Chau, and Bistra Dilkina. Firebird: Predicting fire risk and prioritizing fire inspections in atlanta. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 185–194. ACM, 2016.
11. Michael Madaio, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Bistra Dilkina, and Duen Horng Polo Chau. Identifying and prioritizing fire inspections: a case study of predicting fire risk in atlanta. In *Bloomberg Data for Good Exchange Conference, New York City, NY, USA*, volume 28, 2015.
12. Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. In *International Conference on Learning Representations*, 2018.
13. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014.
14. Cornelis A. J. van Houwelingen and Domien G. M. Beersma. Seasonal changes in 24-h patterns of suicide rates: a study on train suicides in the netherlands. *Journal of Affective Disorders*, 66(2):215–223, Oct 2001.