Abstract: Formula 1 rankings prediction by Neural Networks

September 19, 2018

The objective of this research is to predict the results of a Formula 1 race, because motor racing sports have the greatest number of followers, especially in categories like Formula 1, where the drivers, the new regulations and the incursion of new technologies, give us a sport where prevails the unexpected. One of the activities that most interests fans is to make predictions for each race throughout the season, but there are few technological tools that dare to make a prediction in such a changing sport. This work reviews the most consistent works in similar sports, thats why as initial work it was probed in a neural network with Back Propagation.

The initial work was to collect the data, obtaining the historical information of each race, like who won and the positions of each circuit, so the Ergast [1] database was used. Ergast is an experimental web service that provides motor sports data since 1950 and is updated at the end of each race[1]. Besides it is also free for non-commercial purposes.

The performance of the team is considered relevant, reflecting the positions obtained by its drivers. From these positions is possible to see the points that each driver obtains for his team, which depends on the climate conditions and the race circuit. This variation is due to the fact that a circuit can be favorable to certain equipment depending on the configuration it has, in Formula 1 it is called configuration because a team demands more from the aerodynamic part of the car or if the demand is greater in the engine (more speed). It is also considered by time, since the last races run by a team give us a benchmark of their current performance. So is taken into account the points obtained by each driver for his team, corroborating the data on the official website of the *Federation Internationale de l'Automobile* (FIA) [2].

Having as a result the equation 1, where the team performance DES is defined as the sum of the points obtained by its drivers, divided by the maximum amount of points that is possible to obtain, this helps us to normalize the data.

$$Des = \frac{PO}{MP} \tag{1}$$

where:

PO is the sum of the points obtained by the drivers of the team.

MP is the maximum scores that a team can obtain in a race. This value is 43 being equivalent to the sum of points of the first and second place when they are from the same team.

It is important to note that, in order to obtain data, changes in the names of some teams were considered, so that no information is lost.

The value obtained in the Equation 1 is used for each team, currently there are 10 that will be the number of entry nodes in our network. A multilayered neural network with four layers is used, of which the first is the input layer, the second and third are hidden layers with 32 neurons each, since there is little data and it is necessary for the network to have more operations which it can work with them. And the last one is a layer of 10 neurons representing the drivers who once scored points. Each output neuron has a value between 0 and 1, where the value closest to 1 is the winning driver and a ranking is assigned in a descending order from it. Then, the parameters in table 1 are used:

Network parameters	
Activation function	Sigmoid
Weigth decay	0.01
Learning rate	0.3
Learning rate decay	0.9
Step decay	10
Error	Categorical Cross Entropy

Table 1: Network Parameters

This model has been tested in three races that have different configurations, in the first one, the Canadian GP, where you have the advantage of having a more powerful engine, in the second one, the Italian GP, where more than speed, the aerodynamics of the car matters a lot because it has different types of curves, and in the third one, the British GP, which keeps a balance between speed (engine power) and single-seater aerodynamics (better control in the corners).

Table 2 shows the obtained results.

Pos	PR	P-SGD	
1	Hamilton	**Hamilton**	
2	Vettel	**Vettel**	
3	Bottas	Raikkonen	
4	Raikkonen	Bottas	
5	Ricciardo	**Ricciardo**	
		2	99.4%

Pos	PR	P-SGD	
1	Vettel	**Vettel**	
2	Bottas	Hamilton	
3	Verstappen	Raikkonen	
4	Ricciardo	Bottas	
5	Hamilton	Ricciardo	
		23	93.6%

Pos	PR	P-SGD	
1	Vettel	Hamilton	
2	Hamilton	Vettel	
3	Raikkonen	**Raikkonen**	
4	Bottas	**Bottas**	
5	Ricciardo	**Ricciardo**	
		2	99.4%

Table 2: 2017 Italian Grand Prix, GP of Great Britain 2018 and 2018 Canadian Grand Prix

As can be seen in the table 2 the SGC algorithm had a value of 2, which equals 99.4% accuracy in italian GP. In Great Britain GP the SGC algorithm had a value of 4, which is 92.5% accuracy. And in the Canada GP, the algorithm had a value of 27, which is 92.5%, considering that both *Hamilton*, which was predicted to win the race, and *Raikkonen*, which was predicted to get the third position, had accidents that led them to fall behind of what would be assumed.

With this results is possible to say that neural networks using Back propagation with Conjugated Gradient showed consistent results when predicting actual data with data from the last five years and six last races. The result obtained with this technique lays a solid foundation for the development of this research topic. As future work can be possible to predict how possible is that a team fails, although it is not possible to say how many positions it would affect, it is possible to take into account a possible failure.

References

[1] Ergast. (2018) Ergast developer api. Accedido 07-06-2018. [Online]. Available: http://ergast.com/mrd/

[2] FIA. (2018) Formula one world championship. Accedido 07-06-2018. [Online]. Available: https://www.formula1.com/