Executive Summary Raport 1

Mogage Nicolae – ICA 246/1

Title: Long short-term memory network: A Review

This research focuses on the capabilities of the long short-term memory (LSTM) network in different domains. For a proper analysis, we chose 4 datasets, with 2 being for the classification problem and 2 for the regression. In the first case, we opted for a sentiment analysis prediction dataset, based on twitter messages, with 1.6 million samples. The output expected is whether the text has negative or positive connotations. Next, the second one is an urban sound prediction database with around 8.000 entries of audios, ranging from half a second to 4 seconds sounds. Moreover, these samples cover 10 classes, which are: air conditioner, car horn, children playing, dog bark, drilling, engine idling, gunshot, jackhammer, siren, and street music. Following in the regression case, we went for time-series tabular data. Firstly, a weather forecast dataset was selected which records multiple features, such as pressure, humidity, and wind speed, over a period of years at 10 minutes intervals. The target selected to work is the temperatures in Celsius degrees. Secondly, we wanted to observe the behavior of the LSTM with high variety data, thus we selected a cryptocurrency prediction database with the top 100 coins historically. Out of these 100 possibilities, we went only with 10, having selected: Bitcoin, Cardano, Dogecoin, Elrond (newly EGLD), Ethereum, Litecoin, Polkadot, Solana, Stellar, Tether. The selection was done based on the popularity of the coins and the number of samples they have in order to experience older, but also more recent, cryptocurrencies. These decisions for the datasets allow us to draw the proper conclusions in terms of performance obtained by the LSTM. For the experiments, we split the data intro train, validation and test for the sentiment analysis and weather forecast datasets. For the other two, only train and test fragmentation were used due to the lack of data to work with.

The most important findings suggest that the network performs really good in processing long-term dependencies, but it faces challenges when dealing with shorter sequences. The experiments indicate that the long short-term memory networks are more suitable for regression tasks, especially when predicting continuous variables over time. Moreover, the obtained results prove the capability of the method to capture complex temporal relations in the time-series case if the dataset is large enough and well-processed. Not only this, but it also has an increased performance with high volatile data even if there is less data. On the other hand, the model had a harder time when dealing with short and noisy sequences in the classifications case, failing to distinguish the patterns. For example, the metrics increase very slowly or almost not at all, showing the limitations of the model when deeper contextual understanding is needed. In this manner, further feature engineering techniques or research with other methods that adapt the LSTM networks are recommended to evaluate the implication of this model in more complex situations. Therefore, this research can be seen as an entry point to learn more about this type of network and to form an idea whether is suitable to implement the LSTM in your application or not. It is not excluded that this method is a very powerful one that can lead to accurate, scalable and reliable machine learning solutions.