Project Report

Data Science Fellows February 2020

Galina Blokh

1.Machine-Learning-based Link Fault Detection, TSG IT Systems

2.Improvement of quality of Prediction of Mechanical Ventilation by Decreasing its predictive False Positives, TSG IT Systems

Abstract

1. The purpose of this project was to find a way to predict with a low cost a link fault in a complex network, changing its shift very quickly. The main idea is to be able, by watching the flows in the network and knowing the initial state of the topology, to determine the entrance or the fault of a link in this network by simply monitoring

There were several **main points** to discover in this project:

1. Understand when and where is a fault appearing, in order to deploy an immediate correction
2. Understand how the model is facing the change in the topology due to the integration of a new node
3. Be able to predict the impact of a new node integrating the current topology using Random Forest, XGBoost ML models. Evaluating ML model with ROC curve, Area Under the Curve, Precision, Recall, Confusion matrix
4. Stress test the network on more complex architectures (with 30/60 nodes). The topology can be defined as the blocks with the same structure connected by definite links. The same test as on the small network using Random Forest, XGBoost ML models
5. Test the network on an architecture closer to the client's.

2. Noninvasive ventilation (NIV) may be used in patients suffering from acute respiratory failure. However, failure to deliver the appropriate ventilator support leads to mechanical ventilation which is associated with higher mortality, compared to the patients ventilated from the acute onset. Predicting the success of NIV has been a topic of interest.

mainly varied physiologic markers, including markers for oxygenation and respiratory load. It should be noted that NIV failure may occur shortly after its onset or following hours and even days. Delayed recognition of that NIV failure may lead to increased morbidity and mortality. NIV and High flow Nasal Cannula Oxygenation are used in around 15% of the COVID-19 respiratory failure patients. **Our goal was to improve the prediction of NIV failure** using a newly devised model based on an existing large dataset in respiratory failure.

Challenges

1. The first challenge was in the first “network” project: simulate the network in special application and collect the data. This task (administrating and engineering networks) was built on knowledge which wasn’t provided in ITC study. So quick deepin into a new unknown field was a first challenge.
2. The second challenge was a technical issue, which appeared after 3.5 weeks of work in a “network” project (now the solution is waiting for approval of computer network specialists) and after that relocation to another new project.
3. The third challenge in the new medical project was code, written long long before I came into this project: there weren't docstrings, there were a lot of lines of code and a lot of documents/folders.

Achievements

With network project:

1. Getting new needed knowledge for work in very quick way and short terms
2. Full exploring of technical (key important for the project success) issue and way to solve it for the company

With medical project:

3. Quick deep into new (for me) existing project
4. Using all data science tools for pointed tasks

Further development

On network project:
It was impossible to collect the needed data in given term, thus there wasn't any data science development, but was other related to the project development tasks:

1. Creating real network simulation on Virtual Machine
2. Creating docker image with python script
3. Creating script with using paramico and scapy python libraries to generate simulation of the network flow, collect the data about it, auto-sending it with ssh/sftp connection to the local machine

On the medical project all of this was implemented with account already existing code:

1. Run existing model on validation set
2. To assign true positives as "true" and false positives as "false"
3. To run XGBoost only on the chosen X_Val set and check it on X_Test set
4. To calculate AUC, confusion matrix, all on this new classification
5. Tune XGBoost model for imbalanced data

Supervisor Feedback

Galina did very good work in two tasks that were given to her. Her code is very readable and clear; report and task management were made in a very nice open way also she showed herself as a good team player.

The first task "Machine-Learning-based Link Fault Detection" progressed a lot and now is waiting for additional approval of computer network specialists. Because of absence of time to wait for computer network specialists intervention Galina obtained additional task "Improvement of quality of Prediction of Mechanical Ventilation by Decreasing its predictive False Positives". This task was done with great progress. Now it is in the final phase of the parameters tuning and will be used in the ICU AI computer system installed in Beilinson Hospital.