Paper Report

December 15, 2017

1 Paper: Regresson analysis application

1.1 What problem the paper addresses?

The paper addresses the problem of predicting the energy consumption of a building, solely based on its design parameters. In the very early phases of architectural design of a building, many of its building parameters are not well defined and reluctant to continuous changes, as the designer constantly experiments with different construction approaches. In addition, due to the large number and the complexity of these parameters, the prediction of the designed building enery performance becomes a difficult and time consuming task. In that sense, it is necessary to come up with models that accurately simulate the effect of these parameters, without spending too many computational resources and time to do the prediction. This paper, focuses on identifying a small number of key building parameters that have a significant impact on the energy performance of a building.

1.2 Why is the problem important?

The energy efficiency has become one of the main concerns for new buildings. This comes as no suprice as the energy prices continue to rise and due to the constant effects of the climate change. Furthermore, statistics show that the demand for energy from the commercial sector is projected to increase by 1.2% per year from 2006 to 2030, driven by the population increase and the economic growth. Only in the US the total energy consumption of the buildings corresponds to the 19% of the national total energy use.

1.3 Which datasets are used/tested?

This paper chooses as its baseline model a typicall two-store rectangular office building (i.e modified `ASHRAE 90.1`), which is shown in *Figure 1*. This building has 105 occupants, 1 HVAC system (for cooling or heating) and occupies an area of $2322.6m^2$. More details about this model can be found in the original paper (e.g. net floor-to-ceiling window ratio, floor-to-floor window ratio etc). This building will serve as the simulation model.

However, it is impossible to consider all the possible combination for the design parameters values, due to their huge number. Thus, the *Monte Carlo* method is used which randomly generates samples for each design parameter (variable), based on its original distribution. This way, a uniformly distributed random dataset from available levels of each variable is obtained, which can be fed into the simulation software.

In order to create the dataset, this paper uses the `DOE-2.2` dynamic energy simulation software. This software takes as an input the generated design parameters of the building and produces its estimated total energy consumption. Key parameters relevant to building envelope, such as orientation, occupancy schedules, fenestration, etc. are considered. The climate region of the location of the building is also considered (cold-dry and warn-marine). Moreover, this study tests 8 different configurations of occupant schedules and HVAC operation hours. The results of the `DOE-2.2` simulations are interpreted as the "correct outputs" when feeding the dataset into the linear regression model.

1.4 Which machine learning model is selected to address the problem?

The authors of the paper have selected the multiple linear regression analysis to predict the energy consumption of the buildings. As we have learn in this class the formulation of the multiple-linear regression analysis is the following:

$$Y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

where Y is the total energy consumption, β_j are the regression coefficient and x_i is the $i-th$ simulation vector, whose components are the values of the design parameters. The model was trained with the 80% of the dataset using the *Stepwise regression* method to minimize the sum of squared errors. The other 20% was used for testing and evaluating the model. Because the authors were interested in both the cooling and the heating energy consumption for each building, two regression models were developed for each climate. Thus, in total **four**. The coefficients β_j, which were considered by the regression models (i.e. have their p-value < 0.05), are listed in the Table 1.

1.5 Explain the results of the paper.

One immediate result is that the building that was tested in the cold-dry climate has higher variations in the residual errors compared to the on in the warn-marine climate. This variation can be explained by the distribution of the building energy consumption which is shown in Figure 4. In the first case the points are more concentrated with very few outliers while in the second case the points are more scattered.

Another interesting result is the significant effect of the occupancy schedule to the building total energy consumption. Even if the HVAC system is turned off for two hours more (1h before the working hours and 1h after), this significally affects the total consumption. This holds true even if we have the HVAC system on, one day less.

Besides the occupancy schedule, the exterior wall contruction is another parameter that has a very high impact on the total energy consumption. This was proved by the different types of exterior walls that was used in the two buildings, that was simulated in different climate regions.

Finally, from the 17 building parameters the regression model only kept 14. The type of interior wall, the floor construction material and the type of the roof absorbance were commonly thrown out of the estimator, as they were marked as non significant (and maybe correlated to some other parameters).