Towards the Development of Energy Data Market

Shiuli Subhra Ghosh, MDS202035
Project Supervisor : Dr. Ricardo J Bessa, INESC TEC

September 2021

1 Introduction to Data Market

Recent work has shown that collaboration between data owners may benefit their statistical learning models. However, they may be unwilling to collaborate even if their data privacy is ensured, especially if they do not have relevant benefits. The monetization of data would encourage collaboration between data owners, but the estimation of the data's value and the setting of the corresponding price is a key challenge. Buyers are willing to pay proportionally to the estimation's gains accrued by that marginal data, but such gains are difficult for the seller to infer and therefore to make an attractive price.

The work in [1] proposes an auction mechanism for renewable energy forecasting where all the interactions between data buyers, data sellers and the marketplace itself are considered. The price for the data is defined by the market operator and important properties are ensured:

1. Sellers with similar information will receive similar revenue.
2. The price is relative to the buyer's benefit, and so the buyer does not pay if there is no improvement in the forecasting model.
3. buyers pay according to incremental gain, e.g., if the gain G for bids $b1$ and $b2$ is the same, $G(b1) = G(b2)$ and $b1 > b2$, the buyer payment P is such that $P(b1) = P(b2)$.

2 Project Proposal

Although the previous work addressed several key points, there are still some issues,

1. How to assess a buyer's gain function in other case studies? The map between forecasting errors and monetary evaluation is assumed to be known, however, in many applications this relation is not direct. E.g., how to measure the gain when buying weather forecasts? A possibility is to explore the cost-oriented loss functions such as in [2].
2. How to penalize buyers more efficiently when bid is lower than the market price? The current version works by adding noise to the covariates, which means that, for each new time step, the market operator needs to perform a batch train that can result in a high computational effort as more and more agents enter the market. Ideally, the noise should be introduced in the output of the model, allowing the market operator to update the model weights through online learning whenever the variables in the data market remain the same.
3. Currently sellers' bids are not considered - how to include sellers' bids and design the corresponding pricing algorithm? We could consider diverse privacy sensitivities of data owners. For example, some data owners are less concerned about privacy and are willing to sell their complete sensitive data at a high price, while other providers might not accept disclosure of the complete sensitive data and then they set lower prices but add random noise to their data.

3 Learning Objectives

1. Some new time series forecasting techniques and aligning them with the real-time scenario of the data market.
2. Feature Engineering for numerical weather prediction which is very important for renewable energy forecasting.
3. Mathematical Modelling of Data Market profit functions.
4. Development of new algorithms for payment division and market price update models.
5. Implementing a new incentive scheme for the sellers and buyers.

References

[1] Carla Gonçalves, Pierre Pinson, and Ricardo J. Bessa. "Towards Data Markets in Renewable Energy Forecasting". In: *IEEE Transactions on Sustainable Energy* 12.1 (2021), pp. 533–542. DOI: 10.1109/TSTE.2020.3009615.

[2] Jialun Zhang, Yi Wang, and Gabriela Hug. "Cost-Oriented Load Forecasting". In: (2021). arXiv: 2107.01861 [eess.SY].