CELSE Benoit
IFPEN

Abstract

Machine Learning techniques in Industry. Application to oil refining

In recent years, the increasing integration of the Internet of Things into production industry is at the genesis of a new digital industrial revolution known as Industry 4.0 [1]. The core component of Industry 4.0 is the concept of the digital twin. The main objectives of digital twin are to replace or reduce expensive, time-consuming physical experiments with rapid, inexpensive computer simulation [2]. Accurate and reliable predictive models for physiochemical properties of Hydrocracking process (HCK) products are extremely important. It can help petroleum refinery industries to save time and expansion on costly experiments. In our case, this is the main motivation to synthesize the available knowledge base of HCK process to build a digital twin capable of predicting the product properties of valuable petroleum fractions based on scientific principles. However, for the efficient execution of digital twins is it required to use the different steps of the Knowledge Discovery in Databases-process (KDD-process) [3]. That means automatic extraction of non-obvious, hidden knowledge from large volumes of data [4]. Ideally, the twin would enable:

1. Data cleaning and preprocessing to handle:

   • missing data items by deletion or imputation approach,

   • unexpected values (variables under consideration are expected to have values within a predefined range) by expert pre-processing.

2. Outliers Detection to identify and remove unwanted samples from data. In this work, we use the Local Outlier Factor (LOF) technique [5].

3. Variables selection to select optimum number of variables from a large pool of variables. In the present work we applied leaps [6] and Random Forests [7] algorithms to determine suitable descriptors.

4. Machine learning to build models that characterize the impact of different physicochemical properties of the product. For this, Linear Regression, Kriging ,Support Vector Regression , Random Forest and Gradient Boosting Machine are proposed

5.   Validation of the models using either a dedicated database or cross-validation techniques.

The proposed twin was tested to predict the specific gravity of the vacuum gasoil (VGO), diesel and heavy naphtha cuts in Mild HCK (process that uses low to intermediary pressures and relatively low conversions). The results from this work will be presented. Results are promising but many unforeseen difficulties had to be addressed. In the future, the proposed framework can also be used to predict other properties (cetane number, sulphur and nitrogen content, etc.) with several process (High Pressure Hydrocracking, FCC …). Furthermore, this type of methodology is also extended to predict industrial deactivation.

References

[1]   F. Shrouf, J. Ordieres, G. Miragliotta. Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm. 2014 IEEE International Conference on Industrial Engineering and Engineering Management, 2014, 697-701.
[2]   Grieves M., Vickers J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, in Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Éd. F.-J. Kahlen, S. Flumerfelt, A. Alves. Springer International Publishing, Cham, 2017, 85-113.
[3]   Schuh G., Rudolf S., Riesener M., others. Design for Industrie 4.0. 2016 14th International Design Conference, 1387-1396.
[4]   Fayyad U., Piatetsky-Shapiro G., Smyth P. The KDD Process for Extracting Useful Knowledge from Volumes of Data, Commun. ACM, 1996, 39, 11, 27-34. DOI: 10.1145/240455.240464.
[5]   Ding H., Ding K., Zhang J., Wang Y., Gao L., Li Y., Chen F., Shao Z., Lai W. Local outlier factor-based fault detection and evaluation of photovoltaic system, Solar Energy, 2018, 164, 139-148. DOI: 10.1016/j.solener.2018.01.049.
[6]   Feorge M. Furnival, Robert W. Wilson. Regressions by Leaps and Bounds, 16, 1974.
[7]   Genuer R., Poggi J.-M., Tuleau-Malot C., Villa-Vialaneix N. Random Forests for Big Data, Big Data Research, 2017, 9, 28-46. DOI: 10.1016/j.bdr.2017.07.003.