論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Machine learning-based analysis of solvothermal lignin and lignocellulose liquefaction
著者(和文)	CASTROGARCIA Abraham
Author(English)	Abraham Castro Garcia
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12794号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:CROSS JEFFREY SCOTT,大友 順一郎,髙橋 邦夫,石川 敦之 ,MANZHOS SERGEI,横井 俊之
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12794号, Conferred date:2024/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	融合理工学 系 エネルギーコース コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of (Philosophy)
学生氏名： Student's Name	Abraham Castro Garcia	審査員主査： Chief Examiner	Jeffrey S. Cross

要旨（英文800語程度）
Thesis Summary（approx.800 English Words ）

Thesis title: Machine learning-based analysis of solvothermal lignin and lignocellulose liquefaction

The contents of the thesis chapters are summarized below.

In Chapter 1, the main theme through this dissertation was the usage of ML to clarify and understand the phenomena involved in different kinds of solvothermal lignin depolymerization approaches, the focus being on this particular subset of depolymerization processes due to the authors' perception that it is the most economically feasible and scalable lignin depolymerization process.

Chapter 2 focused on the development of a ML model that aimed to test whether the surface properties of heterogeneous catalysts (surface area, pore diameter and pore volume) played a significant role in the prediction of experimental outcomes from data from heterogeneously catalyzed lignin solvolysis experiments. The results indicate that these properties do not seem to play a major role in the prediction of bio-oil yield, but surface area does predict to some extent the yield of solid residues. For bio-oil yield, process parameters such as temperature, reaction time and choice of solvent seem to play a much larger role than others. This indicates that it is clear that heterogeneous catalysts clearly do play a role in the process, perhaps the majority of the depolymerization that happens is due to the effects of the solvent interacting with lignin at a given temperature for a particular amount of time. It must be noted that the contribution of these heterogeneous catalysts is not only increasing the bio-oil yield, but also changing the product distribution of the process, which is beyond the scope of this dissertation.

Chapter 3 shifted the focus from heterogeneous to homogeneous catalysis, strictly focusing on studies that use water as the solvent of choice (hydrothermal), in attempt to further zoom into how the comparatively few process parameters seen in this kind of process interact. Once again ML models were trained using data from literature obtaining prediction performance similar to that seen in the models in Chapter 2. In this case, unsurprisingly temperature and reaction time were the two process parameters that contributed the most towards the prediction of experimental outcomes, with the catalyst playing a relatively minor role in comparison. This chapter also experimentally tested the predictions made by the models. However, the yield of bio-oil displayed at times an extremely large deviation from the prediction, while the yield of solid residues obtained fell within the expected error range from the model. In the chapter, possible explanations as to why the experimental results for bio-oil deviated so largely from the model's prediction were offered. Possible explanations were offered, first, that the chemistry work-up steps followed after hydrothermal lignin depolymerization may differ across different studies, such as using different solvents to extract a solvent-soluble fraction of lignin from the solid residues, and whether this should or should not be reported as part of the total bio-oil yield (dubbed "Heavy oil" in some papers). Second is that the majority of the studies did not attempt to characterize the lignin that they used in their studies, which likely contributes to the error of the trained models.

Having analyzed the two major groups of studies about lignin solvolysis, Chapter 4 is focused instead in predicting the final HHV of the bio-oil obtained from heterogeneously catalyzed solvolysis of lignocellulosic feedstocks, and also the changes in HHV when said bio-oil is upgraded hydrogenation or thermal treatment. Here once again, processes parameters such as temperature and reaction time proved to be good predictors of the HHV values. Elemental composition, oxygen content in the original feedstock or bio-oil to be upgraded in particular also largely contributed to the prediction. This falls within line of what is currently understood in literature for prediction of HHV from fossil and biofuels, whereas per Dulong's formula, oxygen content in the fuel negatively impacts the resulting HHV, therefore the lower the oxygen content is, the larger the proportion of carbon and hydrogen in the bio-oil, resulting in larger HHV values obtained.

Over the course of this dissertation, it is clear that procuring meaningful amounts of data for experimental work can be a difficult task, and that using data from literature only has clear limitations, such as scarcity and tendency from researchers to report only "good" results, that do not provide useful information to train a ML model. To overcome this weakness, the possibility of miniaturizing and simplifying biomass conversion experiments is put forward, by planning experiments in such way that the results obtained may be representative of the phenomena involved in larger experiments, yet, faster to execute in large quantities. On the other hand, advances in the field of data science are allowing researchers to slowly move away from "big data" approaches that are simply incompatible with practical difficulties seen in experimental work.