Supporting Information for Machine Learning and Text Mining Approaches to Design Selective Catalyst Reduction Synthesis Routes

Shuyuan Li[1]，Chenyu Huang[1]，Yunjiang Zhang[1]，Jing Li[2]，and Shaorui Sun[1]*

[1] Department of Chemical Engineering and Technology，College of Materials Science and Engineering, Beijing University of Technology, Beijing, 100124, China.

[2] School of Mathematics and Physics, Nanjing Institute of Technology, Nanjing 211167, China

*sunsr@bjut.edu.cn

1. Regular expression for catalyst or material preparation paragraph headings.

```
patterns_1 = [
r'catalyst.\w*synth\w*',
r'catalyst.\w*preparation',
r'synthesis.\w*cataly\w*',
r'preparation.\w*cataly\w*',
r'preparation.\w*material\w*',
r'synthesis.\w*material\w*'
]
patterns_2 =
["(?=.*\\bcatalyst\\b)(?=.*\\bpreparation\\b).+",
"(?=.*\\bmaterials\\b)(?=.*\\bpreparation\\b).+",
"(?=.*\\bcatalyst\\b)(?=.*\\bsynthesis\\b).+",
"(?=.*\\bmaterials\\b)(?=.*\\bsynthesis\\b).+",
"(?=.*\\bexperimental\\b)(?=.*\\bprocedure\\b).+",
"(?=.*\\bsynthesis\\b)(?=.*\\bprocedure\\b).+",
"(?=.*\\bsynthesis\\b)(?=.*\\bconditions\\b).+",
"(?=.*\\bpreparation\\b)(?=.*\\bconditions\\b).+"]
```

2. Regular expressions for information extraction
preparation methods:

```
Method_pattern1 =
r"(prepared|synthesized) (by|using|via)(.*?)(technique|method|using|of|\. |\, )\b"
Method_pattern2 =
r'(?i)(impregnation|incipient\s+wetness\s+impregnation|co[-\s]?precipitation|sol[-
\s]?gel|solid[-
\s]?state\s+reaction|chemical\s+vapor\s+deposition|cvd|pyrolysis|microwave[-
\s]?assisted\s+synthesis|ultrasonic[-\s]?assisted\s+synthesis|biological\s+synthesis)'
```

synthetic conditions

```
sentence_dry = 'dried.{0,50}|drying.{0,20}',
sentence_calcine = 'cal-?cin.{0,100}'
sentence_age = 'age.{0,20}'
temperature = '\b\d+(?:\.\d+)?\s*[°°]?\s*[Cc]\b|\b\d+(?:\.\d+)?\s*[Kk]\b|room temp'
time = '\d+.?h|overnight|\d{2,3} ?min|\w{1,7}hours'
```

reaction conditions：

```
GHSV_keywords = ['space velocity', 'GHSV']
GHSV_pattern='(\d{1,4}
?,?\d{3},?\d?\d?\d?)\s*(h−1|h⁻¹|s⁻¹|m/s|mL/(h·g)|mL·g−1·h−1|/h|hr−1)'
temperture_pattern1="\b(\d+\.?\d*)\s*(–|to)?\s*(\d+\.?\d*)?\s*(°C|K)\b(?!\/[a-zA-
Z]+)"
temperature_pattern2 = "\b(ambient|room temperature)\b"
```

performance data:

```
keywors = ['conversion ', 'selectivity']
```

performance_valus = '\d\d.?%|0\.\d{2,4}'

Table S1: Hyperparameter tuning range and parameters setting for machine learning models.

model	Hyperparameter	Parameter

XGBoost	param = {
	"max_depth": [3, 4, 5, 6, 7],
	"n_estimators": [30, 50, 60, 100, 150, 200],
	"min_child_weight": [0.5, 1, 2, 3],
	}
		{'max_depth': 4,
		'min_child_weight': 0
		.5, 'n_estimators': 30
		}

RF	param_rf = {
	"max_depth": [3, 4, 5, 6, 7],
	"n_estimators": [20,30, 50, 60, 100, 150, 200],
	}
		{'max_depth': 5,
		'n_estimators': 100}

Table S2 Model Performance Metrics for Conversion Prediction (Feature: methods, raw materials, synthesis steps, synthesis conditions, and reaction conditions; Raw materials: one-hot encoding)

model	Metric	Train	Test
XGBR	R^2	0.643	0.500
	MSE	0.012	0.012
RF	R^2	0.780	0.757
	MSE	0.007	0.006

Table S3 Model Performance Metrics for Conversion Prediction (Feature: methods, raw materials, synthesis steps, synthesis conditions, and reaction conditions; Raw materials: ASCII encoding)

model	Metric	Train	Test
XGBR	R^2	0.654	0.655
	MSE	0.011	0.009
RF	R^2	0.737	0.728
	MSE	0.009	0.007

Table S4 Model Performance Metrics for Selectivity Prediction (Feature: methods, raw materials, synthesis steps, synthesis conditions, and reaction conditions; Raw materials: one-hot encoding)

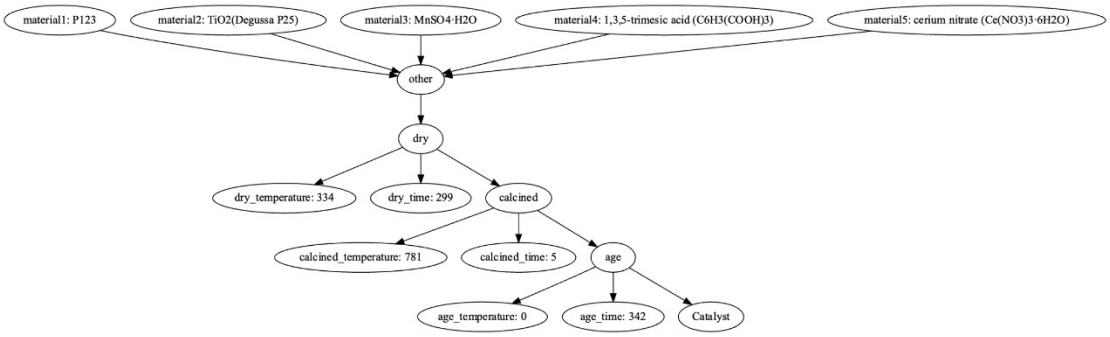
model	Metric	Train	Test
RF	R^2	0.858	0.845
	MSE	0.004	0.010

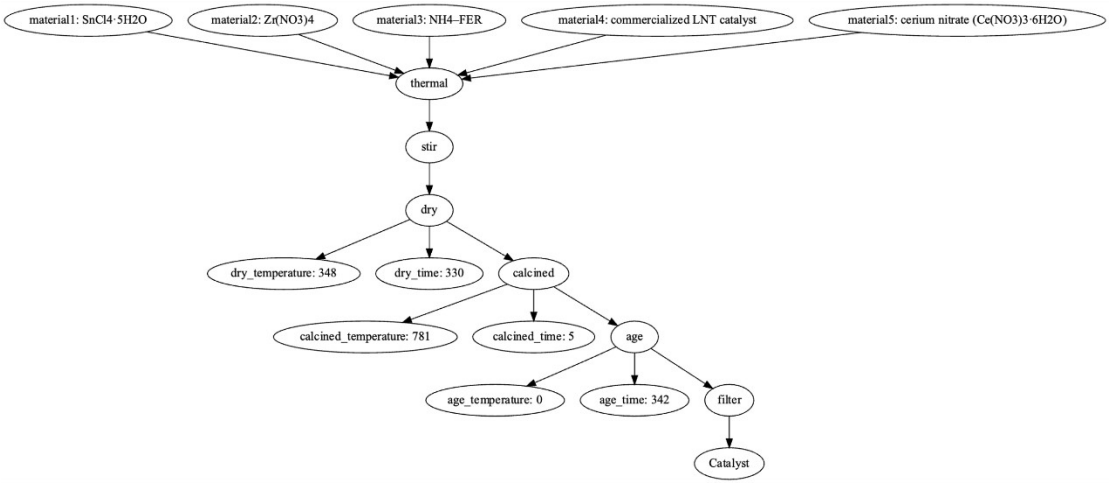
Table S5 Model Performance Metrics for Selectivity Prediction (Feature: methods, raw materials, synthesis steps, synthesis conditions, and reaction conditions; Raw materials: ASCII encoding)

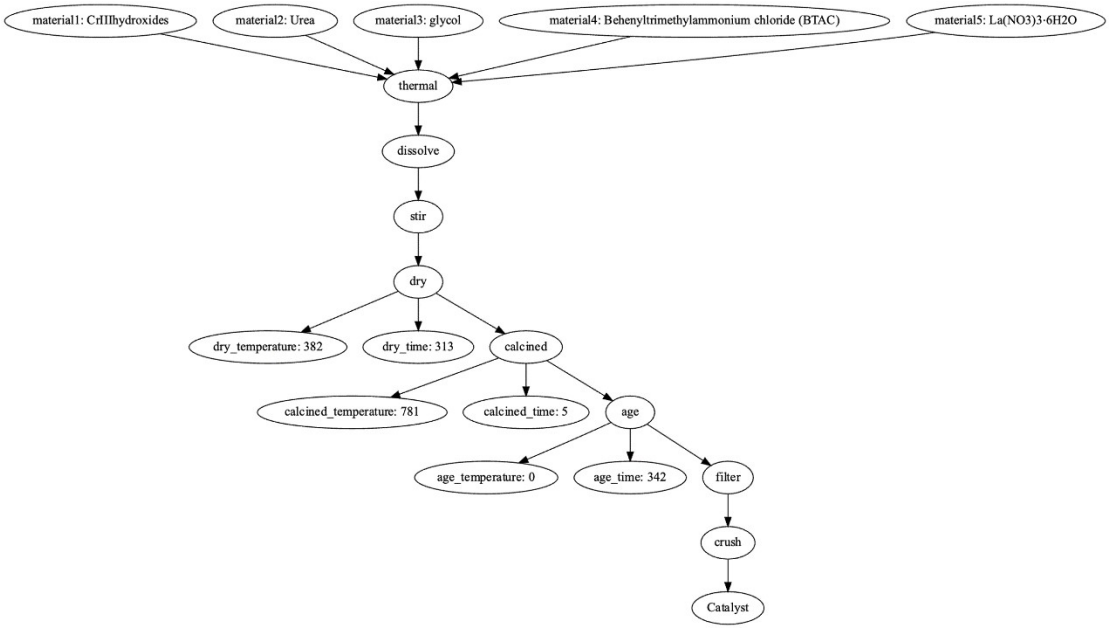
model	Metric	Train	Test
RF	R^2	0.879	0.865

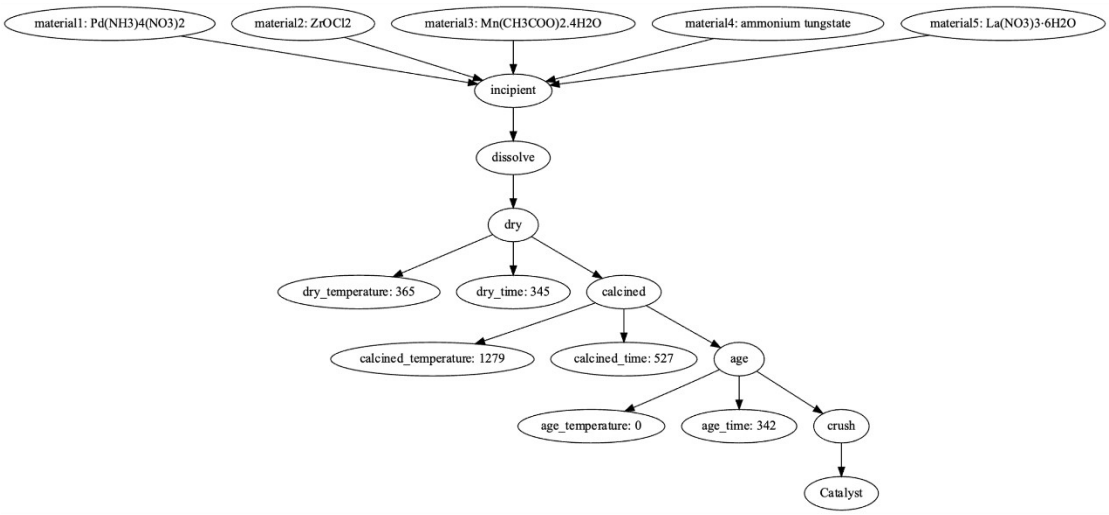
MSE	0.003	0.009

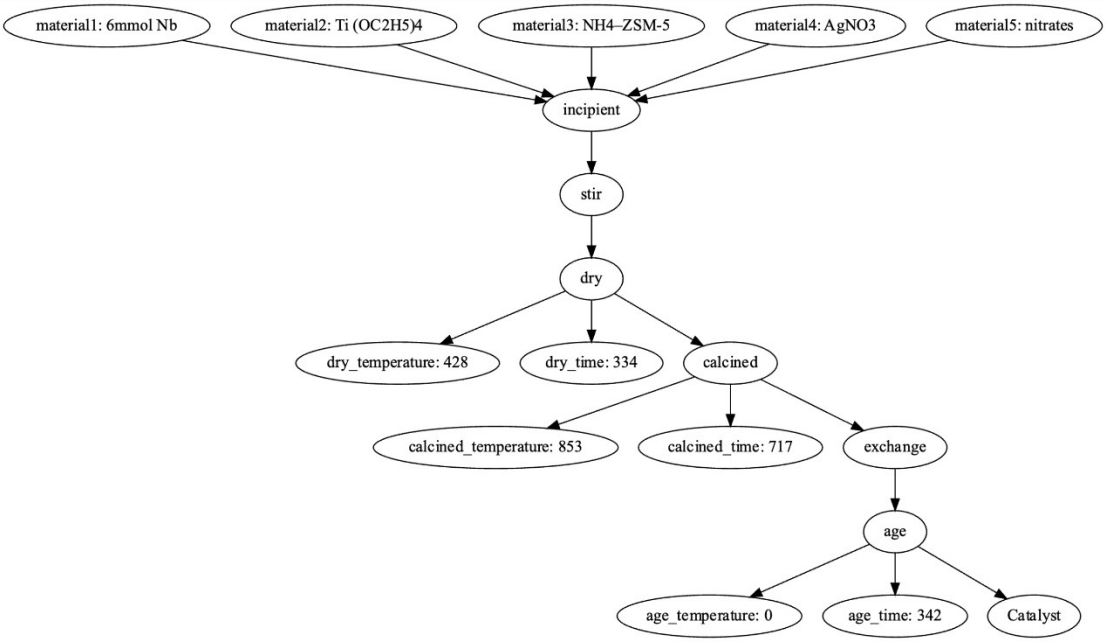
3. Synthetic information in the space was searched and recommended by the machine learning model








material1: Pd(NH3)4(NO3)2
material2: ZrOCl2
material3: Mn(CH3COO)2.4H2O
material4: ammonium tungstate
material5: La(NO3)3·6H2O
incipient
dissolve
dry
dry_temperature: 365
dry_time: 345
calcined
calcined_temperature: 1279
calcined_time: 527
age
age_temperature: 0
age_time: 342
crush
Catalyst


material1: 6mmol Nb
material2: Ti (OC2H5)4
material3: NH4–ZSM-5
material4: AgNO3
material5: nitrates
incipient
stir
dry
dry_temperature: 428
dry_time: 334
calcined
calcined_temperature: 853
calcined_time: 717
exchange
age
age_temperature: 0
age_time: 342
Catalyst