Reaction Chemistry & Engineering

Catalyst Acquisition by Data Science (CADS): A Web-based Catalysts Informatics Platform for Discovering Catalysts

Journal:	*Reaction Chemistry & Engineering*
Manuscript ID	RE-ART-03-2020-000098
Article Type:	Paper
Date Submitted by the Author:	11-Mar-2020
Complete List of Authors:	Fujima, Jun; National Institute for Materials Science, Center for Materials research by Information Tanaka, Yuzuru; National Institute for Materials Science, Center for Materials research by Information Miyazato, Itsuki; Hokkaido Daigaku Kogakubu Daigakuin Kogakuin Kogaku Kenkyuin, Takahashi, Lauren; Freelance Researcher, ; National Institute for Materials Science, Takahashi, Keisuke; Hokkaido University, Department of Chemistry; National Institute for Materials Science, Center for Materials research by Information

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Catalyst Acquisition by Data Science (CADS): A Web-based Catalysts Informatics Platform for Discovering Catalysts†

Jun Fujima*[a], Yuzuru Tanaka[a], Itsuki Miyazato[ab], Lauren Takahashi[ab], Keisuke Takahashi,*[ab]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

An innovative web-based integrated catalysts informatics platform, Catalyst Acquisition by Data Science (CADS), is developed for use towards the discovery and design of catalysts. The platform provides three main functionalities: a repository for data sharing and publishing, an analytic workspaces for exploratory visual analysis, and catalyst property prediction tools with pretrained machine learning models. Access to such a platform helps decrease barriers to entry faced by researchers in catalytic chemistry when attempting to apply catalyst informatics towards data by providing analytical and visualization tools that can be simultaneously applied and easily accessed within a central space, thereby helping the advancement of catalyst informatics. The developed platform allows researchers to upload and collect data onto the platform and conduct data analysis using a system of linked workspaces consisting of interactive visualization tools and machine learning tools that simulataneously update according to the researchers' actions in real time. The platform also provides a space for collaboration where researchers can choose to publish their uploaded data and resulting analyses to the platform for collaborations with other users and groups. As an example, CADS is applied towards oxidative coupling of methane (OCM) data where use of the platform tools reveal underlying patterns and trends that were otherwise hidden within the original data. Thus, the proposed platform contributes towards the advancement of catalysts informatics for both specialists and non-specialists.

Introduction

The implementation of data science techniques has greatly empowered catalyst data thanks to the rapid growth of data available for catalysis research[1–3]. This development is referred to as catalysts informatics, where design and insight into catalysis arises from the trends and patterns found within catalyst data through data science applications[4,5]. However, while it shows much promise, catalyst informatics involves factors such as data construction, data management, visualization, machine learning, and a variety of other data analysis skills, thereby requiring a wide range of data science techniques and related knowledge[4,6]. Furthermore, these tasks are not often collected or developed together, making it difficult to link data construction to catalyst design. Such individualistic development has limited the ability to properly implement catalyst informatics, making it even more necessary to have a centralized, standard platform available for catalyst informatics applications.

Similar issues have been previously faced in the early stages of bioinformatics and materials informatics where web-based platforms providing analysis for informatics have been developed for solving such issues[7,8]. Within the scope of materials informatics, various databases and platforms consisting of data generated from first principles calculations as well as data and code related to computational materials science are available for use with examples ranging from the Open Quantum Materials Database (OQMD), the Novel Materials Discovery repository (NoMad), and Automatic Flow for Materials Discovery (AFLOW)[9–11].

Along these lines, web-based catalysis data platforms have been also developed with examples such as CatApp and Catalyst Hub[12,13]. Unfortunately, these platforms lack data analysis functions and components, and more closely resemble data search engines with available visualization functions. Additionally, it is very difficult for experimental catalyst researchers that are unfamiliar with data management or software programming to use

[a] Center for Materials research by Information Integration (CMI²), National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan
* email:jun.fujima@nims.go.jp, keisuke.takahashi@eng.hokudai.ac.jp
[b] Department of Chemistry, Hokkaido University, Sapporo 060-8510, Japan.

data provided by such data plaforms. While some of the platforms provide data visualization functionalities, they are often very limited to a specific usage scenario. Hence, it is crucial to develop a platform that is accessible and is equipped with the tools necessary for catalysts informatics.

The core concept for a catalyst platform is to provide an environment that can not only be easily accessible but can also simultaneously handle data management and data science techniques via a graphic user interface. Here, a web-based interactive platform named Catalyst Acquisition by Data Science (CADS) is proposed where the process of data construction to catalyst design can all be achieved in one location where the proposed platform is available at `https://cads.eng.hokudai.ac.jp/`. In particular, through using the platform researchers can visualize and analyze data simultaneously, where data analysis visualizations, tables, and other tools like machine learning methods are linked and updated in real-time, making the process very interactive and suitable for applying data science towards catalysis data analysis. Furthermore, researchers can publish catalyst data as well as resulting analyses and share with other researchers for collaboration efforts. The introduced platform thus contributes towards the advancement of catalysts informatics for both specialists and non-specialists.

Concept and Platform Overview

Integrated Platform for Catalysts Informatics

There are numerous resources available for catalayst informatics applications, though they are available in an uncoordinated manner. To start, experimental scientists collect catalyst data using a variety of different approach through their research activities. Simultaneously, developments in high-throughput experimentation and calculations have accelerated the accumulation of catalyst data both qualitatively and quantitatively. In order to apply data science techniques towards data generated and collected through these methods, it becomes necessary to translate the data into its digital equivalent, organize said data in a uniform manner, and store the results in a shared repository.

On the other hand, the advancement of internet and computing technology had made it possible to share and reuse various types of data visualization and data analysis methods, including machine learning, as software libraries and frameworks. In comparison, skilled researchers can combine these libraries and frameworks with relative ease in order to analyse data. However, doing so requires the knowledge of some programming languages (such as Python) as well as the skill and knowledge of how to access and utilize these libraries and frameworks. As many researchers do not have these skills, there is a large barrier in place for non-specialized researchers that may wish to apply data science towards their research. Furthermore, even if one were posses such skills, analysis of heterogeneous data like catalyst data requires a trial-and-error approach in order to determine the most relevant descriptors of the given datasets as well as the analysis or machine learning methods that would be most appropriate for that particular data set. This results in very tedious work even for skilled researchers as it involves repetitive changes to programming scripts

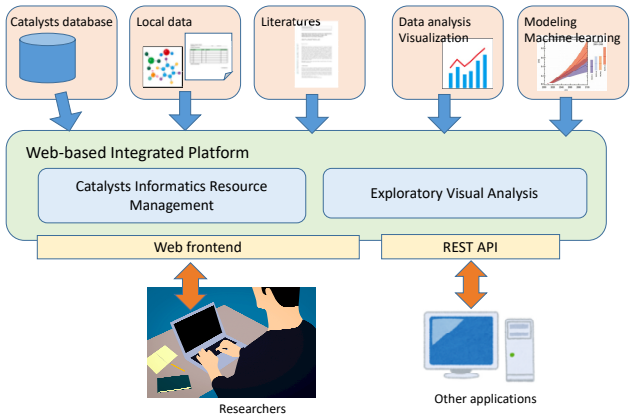


Fig. 1 The concept of the proposed platform.

when investigating different analysis settings and evalutation of the data analysis results.

In order to help propel the development of catalyst informatics forward, it becomes necessary to provide an integrated platform where researchers can manage resources for catalyst informatics applications in a central location. Figure 1 depicts the basic structure and schemes of the proposed platform. In such a platform, researchers must be able to accumulate catalyst data and to successfully utilize the offered analysis and visualization techniques, including machine learning techniques, without requiring additional knowledge relating to computer programming or machine learning. All necessary resources including datasets, analysis methods, visualizations, and machine learning techniques should also be published and made accessible through the platform in an unified manner so that other researchers or other programs can reuse such resources for further analysis. Considering these factors, the developed platform aims to provide these functionalities in an integrated environment that is easily accessible to any catalyst researcher.

Platform Overview

An innovative platform for a catalyst informatics platform, Catalyst Acquisition by Data Science (CADS), is developed in efforts to address the needs of catalyst researchers that wish to apply informatics towards their research. The proposed platform provides three sub applications: *Data Management*, *Analysis*, and *Prediction*. The platform allows researchers to upload local files containing catalyst data in the *Data Management* application. Once a file is uploaded upon the platform, researchers can then analyze the underlying data in the *Analysis* application where various analysis methods are made available on the platform. These methods include not only simple visualization tools such as scatter plots but also complex analysis tools such as regression and clustering methods used in machine learning. Using trained models created in the *Analysis* application, researchers can then attempt to predict particular physical properties of catalysts.

Interested readers may refer to the documentation of the CADS

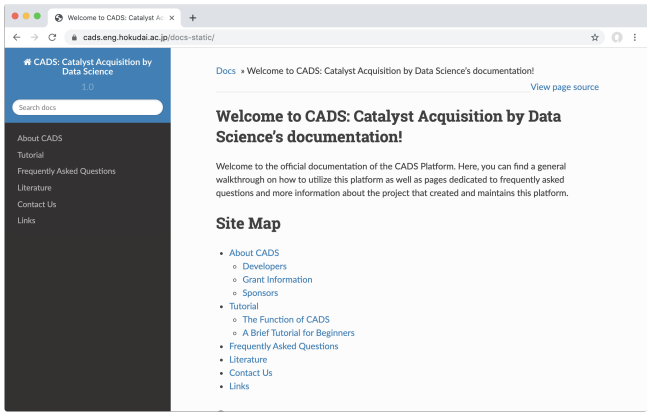


Fig. 2 The documentation page of the Catalyst Acquisition by Data Science (CADS) platform.

platform (Figure 2 *). The documentation provides further description and general walkthrough tutorial on how to utilize the platform.

Data Management

Researchers (which, from this point, are also referred to as "users") can upload arbitrary files that contain catalyst data in table form. Once a file is uploaded to the application, the file is recognized as a *data source* within the platform. Currently, the platform supports files formatted as Comma Separated Value (CSV) files for using the data in the *Analysis* application.

Figure 3 shows a snapshot of the *data management* application. When a user accesses the *data management* application, all available data sources that have been previously uploaded to the system or shared by other users are listed at the top page and are searchable by keywords. Users may select a data source in order to browse the content of a data source, which is displayed in tabular form. The displayed table can be narrowed down through the use of search keywords while specific columns may be made visible or hidden at the choice of the user.

Users also have control over the accessibility of the data sources. There are three levels of accessibility on the platform: *private*, *internal*, and *public*. The default setting at the initial creation of a data source is *private*, where only the file owner can read, use, or delete the data source in question. The owner of the data source also has the ability to change the permissions to other levels of accessibility at their discretion. For instance, data sources marked as *internal* may be accessed by users with the proper permissions while data sources marked *public* can be accessed by all users. Shared data sources are published using copyright licenses provided by the Creative Commons †.

Currently, the platform publishes four public data sources: "ChemCatChem" data (consisting of catalyst data in relation to the oxidative coupling of methane reaction), "Perovskite" data (consisting of data relatigng to a series of materials with a cubic perovskite structure), "2D Materials" (consisting of a variety

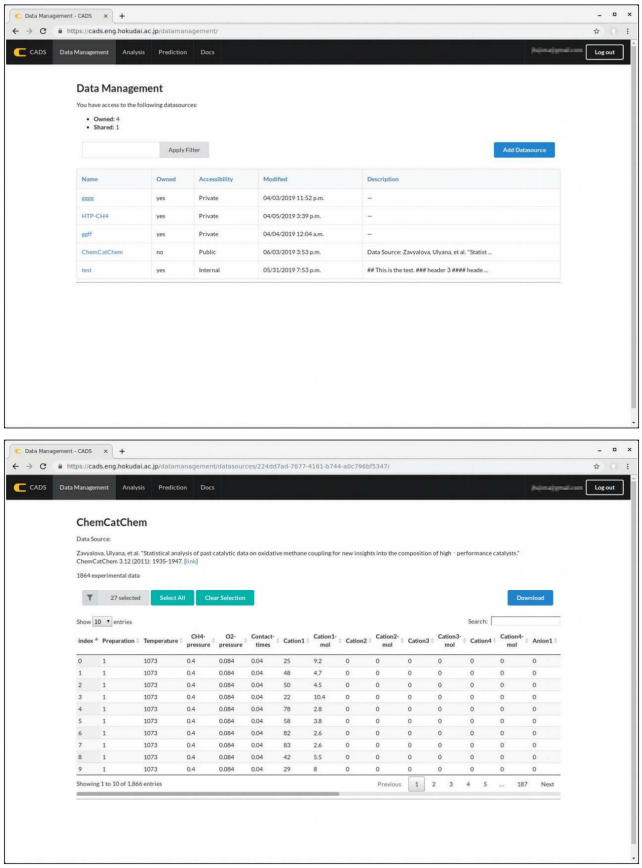


Fig. 3 *Data Management* sub application: a list of uploaded data sources on the top and a table view of the content of a data source on the bottom.

∗ https://cads.eng.hokudai.ac.jp/docs-static/
† https://creativecommons.org/

of data relating to approximately 1500 two-dimensional materials), and "CatApp" data (consisting of data relating to activation and reaction energies calculated for coupling reactions on metal surfaces) [12,16–18]. Note that the Perovskite, 2D Materials, and CatApp databases are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Analysis

The *analysis* application houses the core features of the developed platform. In order to provide users with a flexible and easy-to-use environment for analysis of heterogeneous data, coordinated multiple views (CMVs) are employed [19–21]. Coordinated multiple views is an exploratory visualization technique that combines multiple *views*. Utilizing multiple views allows for the visualization of target data from multiple viewpoints simultaneously. Each view displays the target data with a different visualization technique and allows users to select a set of visualized objects through direct manipulation. This selection defines a new condition for highlighting the corresponding objects and is immediately reflected within the other views. Users are also able to repetitively set up different selections in order to investigate how different conditions are affected through different visualizations. Sjöbergh et. al. have previously extended the original coordinated multiple views so that statistical analysis and data mining tools are available as views within their CMV framework [22,23]. In light of this, a visual analysis framework is developed with various functionalities that aids users when attempting to apply data science techniques, including machine learning, towards catalysts informatics research.

Figure 4 illustrates an analysis result generated through the use of the *analysis* sub application in the developed platform. A user first selects a target data source for analysis from the stored data sources made available to the user. The user can then apply different views in order to visualize the selected data from different view points. As seen in Figure 4, the user has applied six views: two scatter plots with different settings, a table view, a feature importance analysis tool, a regression tool, and a histogram tool. Additionally, users also have the ability to save the workspace containing the analysis results. Similarly to data sources, one can assign accessibility to saved workspaces in order to control the access levels of the workspace.

Users may also save trained machine learning models such as regression models for future reuse within the *predict* sub application. This ability allows users to calculate estimated property values for given input conditions.

In section 3, we describe the implementation of the Coordinated Multiple View framework in further detail.

Prediction

Users can estimate a specific physical property corresponding to a particular set of parameter values using stored pre-training machine learning models. Figure 5 presents a snapshot from the *Prediction* application. Users can enter explanatory variables into the input fields to specify desired conditions for estimated physical properties of catalysts. Then, pressing the 'Predict' button triggers a calculation of the estimation. The result of the calculated estimation is displayed at the bottom of the web page.

Implementation

System Architecture

Figure 6 depicts the architecture overview of the proposed platform. The introduced platform is implemented as a web application. Uploaded data sources are stored in the data store as file objects. As the application database backend, the Postgres relational database is used, which manages application-related data such as resource metadata or user information. Users can access the application through the reverse proxy.

Python is implemented as the main programming language of the platform to allow data analysis to access machine learning libraries such as pandas [‡] or scikit-learn [§] while Django is implemented as the base framework for server-side functionalities of the main web application and web application programming interfaces (APIs). The Javascript framework React [¶] is used client-side in order to share parts of the display functionalities and interactive operations. For realizing interactive graph displays, Bokeh is implemented [‖].

Visual Analysis Framework

A framework for visual analysis is implemented where coordinated multiple views are utilized. This type of framework allows users to combine multiple view components in order to analyze catalysis data on demand.

Figure 7 illustrates an overview of the framework and how the components communicate with each other. The main components of the proposed framework are Views and a ViewCoordinator. The ViewCoordinator opens and manages a workspace that coordinates all views added by the user. When the user selects a data source, the ViewCoordinator fetches the content of the selected data source and holds it in a tabular form. This table is used in order to coordinate different views while the workspace component manages the visual arrangements of the views. By selecting the 'Add view' button on the workspace, the user may add new views in order to set up new visualizations or analysis results at their discretion.

Views

A view is a visual wrapper component of an arbitrary analysis or visualization function. It provides a display which represents data with specific visualization technique as well as a common interface for controlling the behavior of the corresponding view function. Figure 8 presents a snapshot of a scatter plot view.

As the common user interface, a view has a 'Close' button for removing the view from the workspace and a 'Setting' button to open the setting panel of the view. With the setting panel, users can specify parameters of the view function. In the example of a scatter plot, users can specify the row name of the target data

‡ https://pandas.pydata.org/
§ https://scikit-learn.org/
¶ https://reactjs.org/
‖ https://bokeh.pydata.org/en/latest/

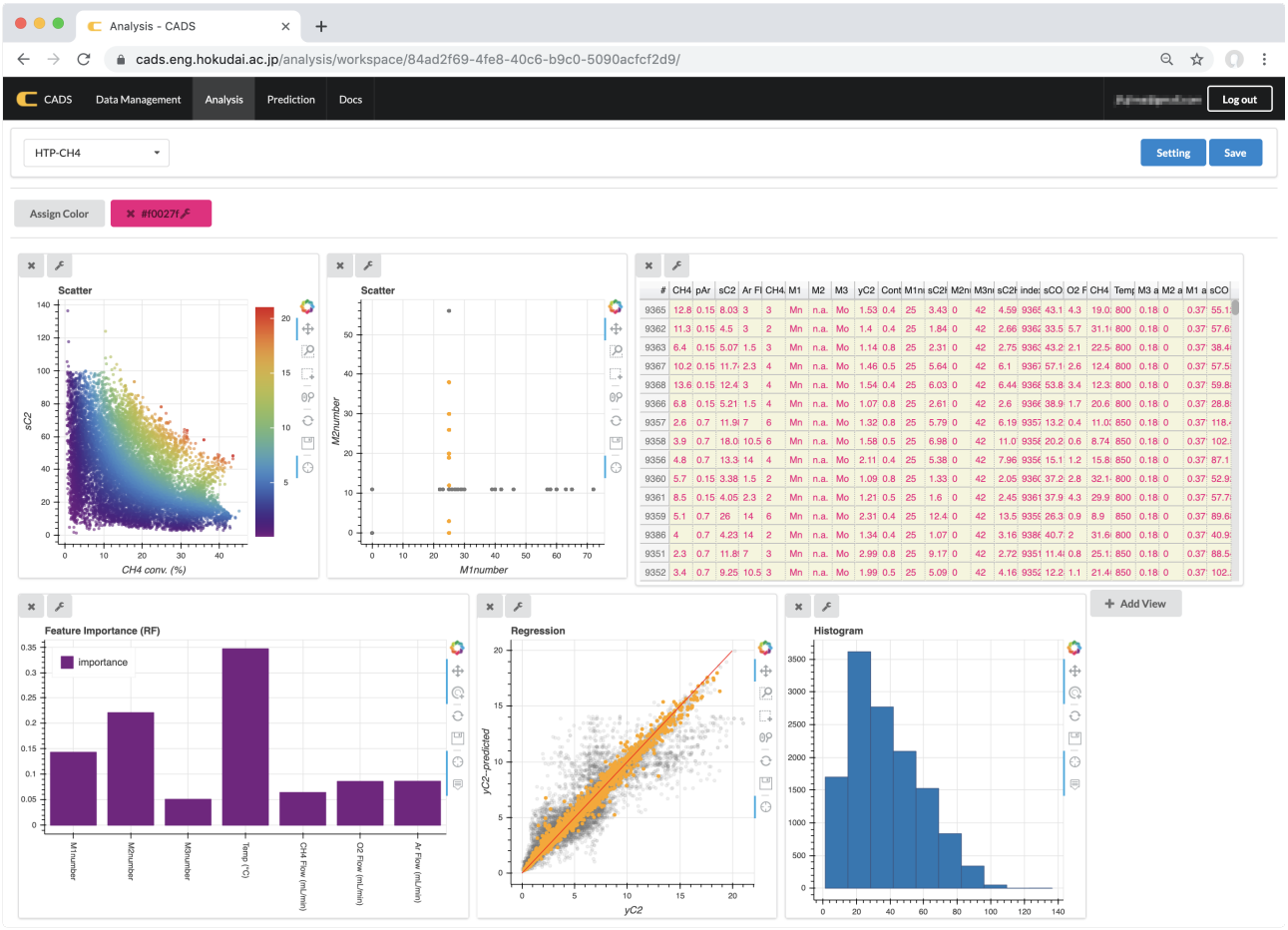


Fig. 4 *Analysis* sub application with coordinated multiple views consisting of six views.

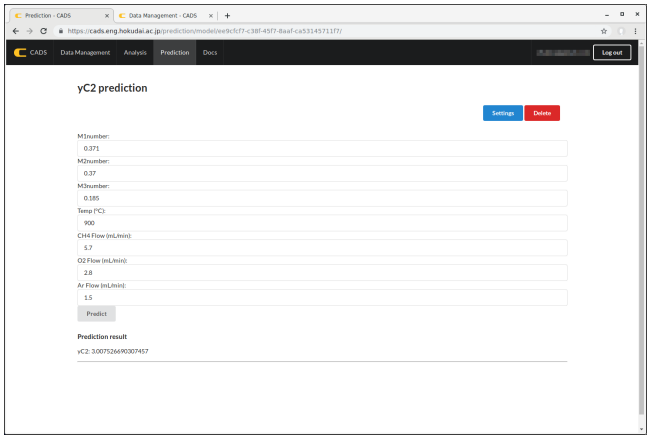


Fig. 5 *Prediction* sub application where users can make predictions with pre-trained models.

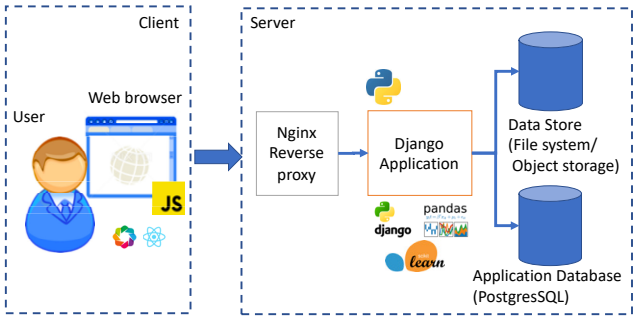


Fig. 6 System architecture of the proposed platform.

source assigned for the X axis and Y axis and also configure the color assignment of the display objects in the plot.

A view function can vary from a simple view that only visualizes the original data as a plot to a view that involves complex calculations like machine learning. Some complex components access the servers with the REST protocol for making heavy calculations on the server side and receiving results to make visualizations of the result data.

Currently, the provided views in the platform are categorized into the following three categories:

1. visualization:
   scatter, table

2. analysis:
   histogram, clustering, feature importance with random forest,

3. machine learning:
   regression, classification.

Views in the visualization category are simple visualization components that assign data values with visual properties where all necessary calculations are carried out from the client side. On the other hand, views from the analysis and machine learning categories involve more complicated calculations and use the REST APIs to create visualizations of the results.

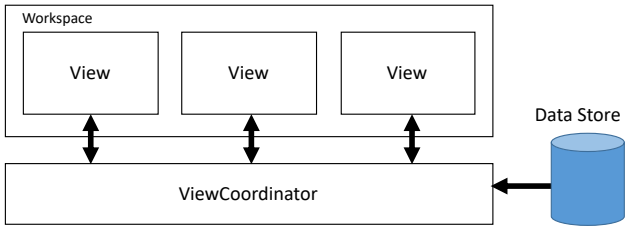


Fig. 7 The overview of the visual analysis framework.

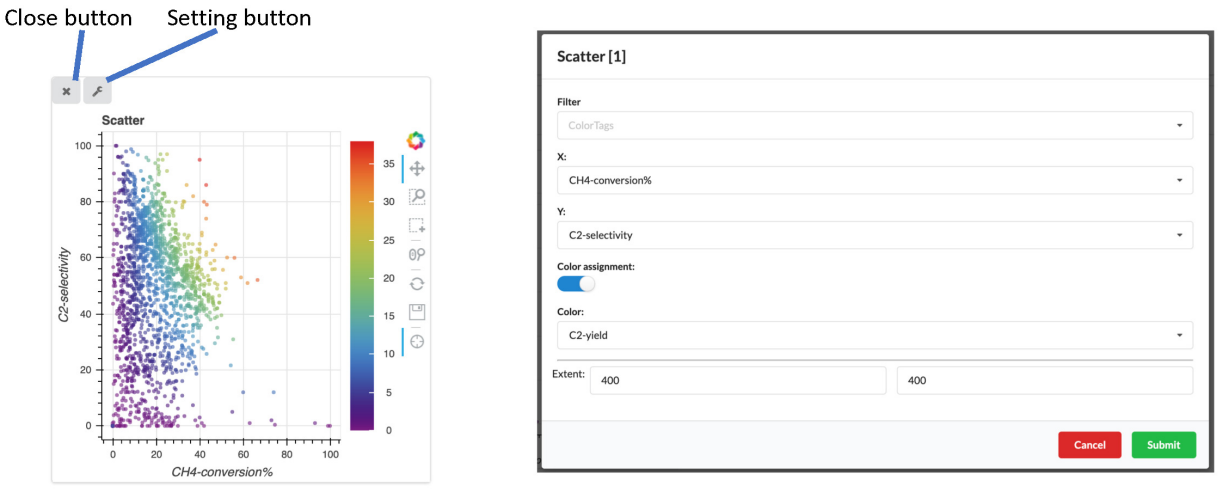
Data Linkage between Views

Views are connected and interact with eachother through the ViewCoordinator. The ViewCoordinator fetches data from a selected data source and notifies the views of the change in target data. The views then update their display according to their settings upon incorporating the new target data. When the user selects a set of data objects in a view, the selection operation is transferred to the ViewCoordinator. The ViewCoordinator then notifies all views in the workspace of the selection change. As a result, the views update the selection within their display so that the data is in the same row on the internal table as with the selected data object. In this way, the brushing operation is realized among different views.

Application

The power of CADS is demonstrated using OCM data where OCM data consisting of 1866 data is collected from literatures[1]. One of the main features of CADS is interactive visualization which allows users to reveal the patterns and trends hidden within data in an interactive manner. Two scatter plots, shown in Figure 9, are visualized using 1866 OCM data where C_2 selectivity vs CH_4 conversion and CH_4 pressure vs O_2 pressure are plotted. Here, three exploratory data analyses are interactively carried out in order to reveal the relation of how C_2 selectivity and CH_4 conversion behave against CH_4 and O_2 pressure. The orange plot points represent the selected data points within the dataset chosen for analysis. As can be seen in the graphs, the selected data points within the dataset are highlighted and are linked to each other. This allows one to select data points in one graph and have that selection be automatically highlighted in the linked windows. This feature allows for simultaneous analysis of data across multiple windows.

In the case of Figure 9, high CH_4 conversion and low C_2 selectivity are achieved when high O_2 pressure is applied. One can consider that large number of oxygen is responsible for oxidation of CH_4, resulting the high CH_4 conversion. In same manner, how CH_4 pressure affect the C_2 selectivity and CH_4 conversion is investigated. Figure 9 shows that high CH_4 pressure results in low CH_4 conversion and high C_2 selectivity while low CH_4 pressure results in relatively high CH_4 conversion. These observations and conclusions were able to occur thanks to the ability to link windows, which allow them to interact with each other. As a result, interactively selecting the multiple scatter plots allows for the revelation of hidden information. Hence, interactive exploratory data analysis in CADS can assist the design of catalysis experiments.



(a) The common interface of a view (Scatter plot) (b) The setting panel of the scatter plot

Fig. 8 The common interface of a view for the case of a scatter plot.

Machine learning function in CADS is also explored using the 1866 OCM data. Here, random forest regressor(RFR) is implemented where the objective variable is set to C_2 yield. The following 6 descriptor variables are chosen for training RFR: Cation1, Cation2, Contact-times, CH4-pressure, O2-pressure, and Temperature while C_2 yield is . Figure 10 represents the scatter plot of C_2 selectivity vs CH_4 conversion as well as true and predicted C_2 yield via RFR. Note that the presented predicted and true C_2 yields in Figure 10 are the result of only trained data for demonstration purposes; therefore, prediction ability towards test data is not evaluated here. Overall, Figure 10 shows that prediction of C_2 yield is good although there are a few points of predicted C_2 yield that are not predicted well. Here, CADS plays a major role in revealing which data points are not being predicted. In Figure 10, data points which are not accurately predicted in RFR are selected, where the corresponding data points are then automatically highlighted in scatter points. It is interesting to see that the data points that are not accurately predicted in RFR are the data points which have high C_2 yield: more specifically, C_2 yield over 30%. This suggests that C_2 yield above 30% has different patterns and rules in the OCM data in comparison to yields less than 30%. Thus, CADS can interactively unveil the data points that do not follow the general trends in rule, thereby demonstrating that CADS can provide guidance for consecutive data analysis as well as scientific analysis.

Lastly, OCM data via high throughput experiments are visualized using the CADS platform. The high throughput OCM data consists of 12,706 data[28]. Here, the following seven scatter plots are visualized in Figure 11: C2s vs CH4cov, Arflow vs Temp, O2flow vs CH4flow, COs vs CO2s, C2H4s vs C2H6s, M1atomnumber vs SupportID, and M2Atomnumber vs M3atomnumber. Please see the reference[28] for the details of each variable.

In Figure 11, the data points with high C_2 yield (in other words, high C2 selectivity and high CH_4 conversion) are selected and highlighted throughout various scatterplots with different variables. Immediately, one can see that there are trends present with data points that have a high C2 selectivity and high CH_4 conversion. For instance, it becomes clear that such data points tend to occur when temperature is between 700C and 850C where the upper end of this temperature range results in high C2 selectivity when Ar flow is increased. Additionally, while there is no particular trend when comparing CH_4 flow and O_2 flow, there is a trend present when considering CO and CO_2 selectivities. In particular, the selected data points are found to concentrate where CO and CO_2 selectivities are low, which suggests that suppressing CO and CO_2 production is important. The selected data points are also found to result in large C_2H_4 selectivity (ranging between 30 to 70) where C_2H_6 selectivity increases as C_2H_4 selectivity increases, suggesting a correlation between both selectivities. Finally, it also becomes possible to understand which combinations of atomic elements result in high rates of C2 yield by plotting atomic numbers of each element against its appropriate support ID (which represents the chemical formula being considered). In comparing these variables, trends in terms of atomic composition become clearer to understand. For instance, when the atomic number of element 1 is 25, the choice of support can affect C2 yield rate. Additionally, when comparing the second and third elements of the catalyst, it becomes clearer that high C2 yield is likely to occur when the second atomic number is 11 or 19 and the third atomic number is either around 42 or 74, suggesting that atomic composition is also important to consider when looking for catalysts that result in a high C2 yield. These results thus demonstrate the power of the CADS platform for multi-dimensional analysis for data that allow for the discovery of underlying trends within data where the platform's ability to simultaneously plot and illustrate the relationships of various data and variables allows one to compare data points on a large scale that is both convenient and easy to analyze.

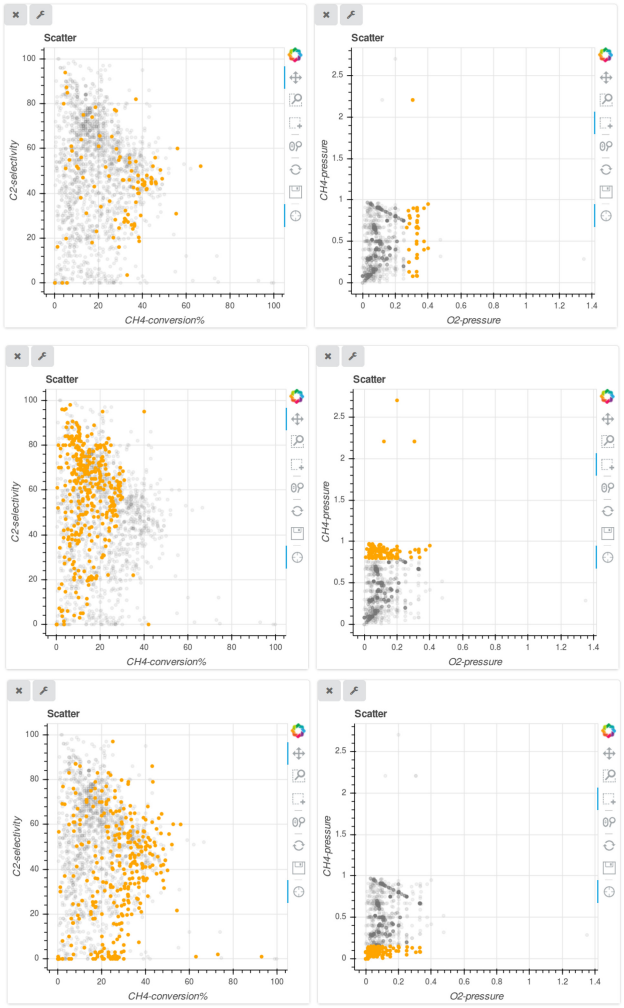


Fig. 9 Scatter plots of C_2 selectivity vs CH_4 conversion and CH_4 pressure vs O_2 pressure.

Conclusion

A catalyst informatics platform, Catalyst Acquisition by Data Science (CADS), is developed and proposed. The proposed platform provides a data management system for publishing and sharing catalysts data among users. It also provides an interactive visual analysis environment for exploration of uploaded catalysts data. Such an interactive environment assists researchers in discovering trends and patterns hidden within catalyst data. In applications of CADS, OCM data is analyzed where the reactions between CH_4 and C_2 pressures against C_2 selectivity vs CH_4 conversion are explored. Furthermore, implementation of RFR in CADS reveals the hidden patterns in OCM data set. Thus, CADS provides a multi-functional environment for catalyst informatics that can assist researchers in designing catalysts from catalysts data. Lastly, further functions in CADS are expected to be developed as well as to help encourage researchers to share catalysts data for the community.

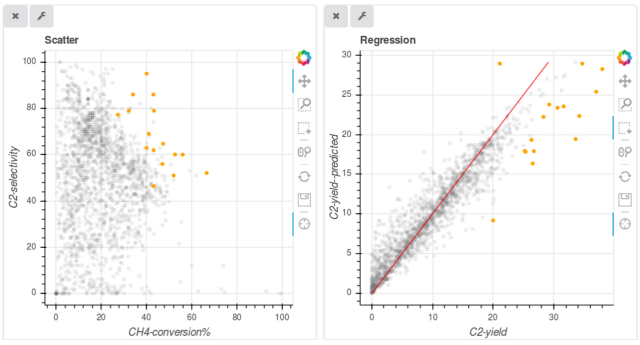


Fig. 10 Scatter plot of C_2 selectivity vs CH_4 conversion and predicted and true C_2 yield.

Acknowledgment

This work is funded by Japan Science and Technology Agency (JST) CREST Grant Number JPMJCR17P2, JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Grant Number JP19K11988, JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Grant Number JP17K14803, and Materials research by Information Integration (MI^2I) Initiative project of the Support Program for Starting Up Innovation Hub from JST.

Notes and references

1 U. Zavyalova, M. Holena, R. Schlögl, and M. Baerns, ChemCatChem **3**, 1935 (2011).
2 J. K. Nørskov and T. Bligaard, Angew. Chem. Int. Ed. **52**, 776 (2013).
3 Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, Nat. comm. **8**, 14621 (2017).
4 K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, and K. Hirai, ChemCatChem **11**, 1146 (2019).
5 P. Schlexer Lamoureux, K. Winther, J. A. Garrido Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen, and T. Bligaard, ChemCatChem (2019).
6 A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, and R. Fushimi, ACS Catal. **8**, 7403 (2018).
7 Y. Yu, M. Breitbart, P. McNairnie, and F. Rohwer, BMC Bioinform. **7**, 57 (2006).
8 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, and G. Ceder, APL Mater. **1**, 011002 (2013).
9 S. Kirklin, J. Saal, B. Meredig, A. Thompson, J. Doak, M. Aykol, S. Rühl, and C. Wolverton, Npj Comput. Mater. **1** (2015), 10.1038/npjcompumats.2015.10.
10 C. Draxl and M. Scheffler, MRS Bull. **43**, 676 (2018).
11 S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, and O. Levy, Comput. Mater. Sci. **58**, 218 (2012).
12 J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard, and J. K. Nørskov, Angew. Chem. Int. Ed. **51**, 272 (2012).
13 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich, and T. Bligaard, Sci. Data **6**, 75 (2019).

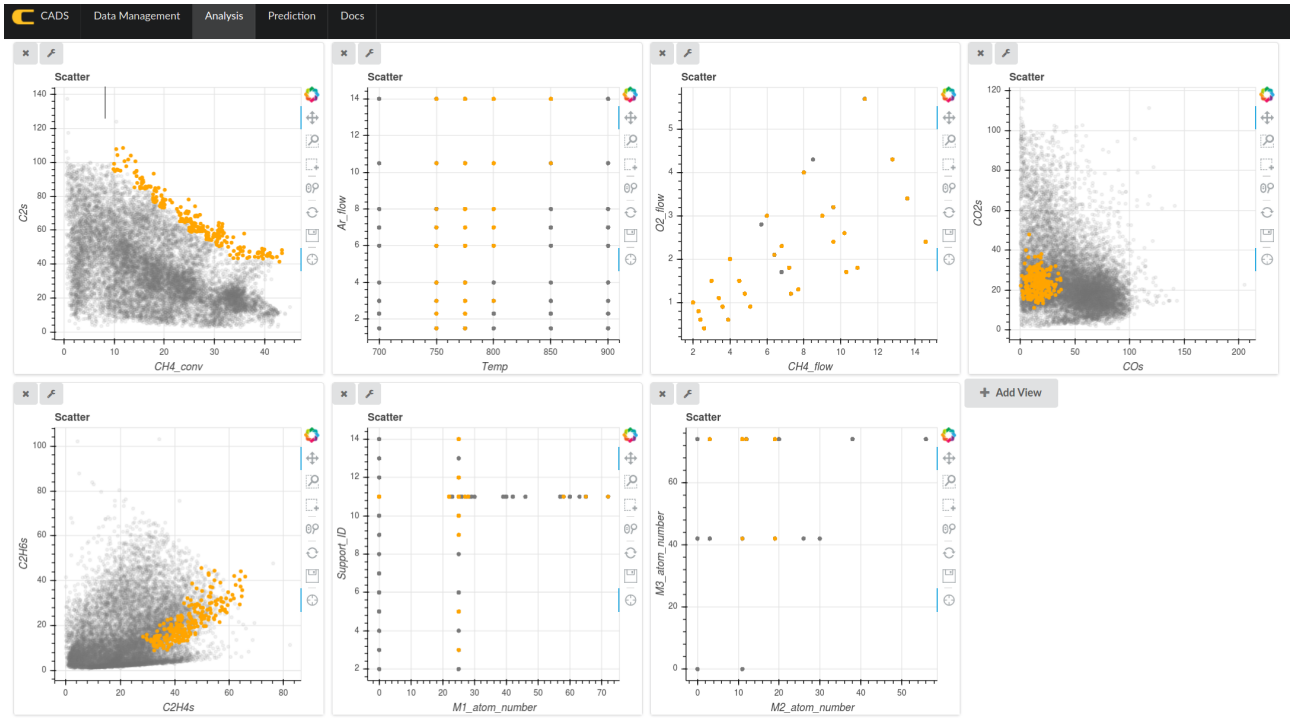


Fig. 11 Scatter plot of HTP OCM data[28].

14 https://cads.eng.hokudai.ac.jp/docs-static/.

15 https://creativecommons.org/.

16 U. Zavyalova, M. Holena, R. Schlögl, and M. Baerns, ChemCatChem **3**, 1935 (2011), https://onlinelibrary.wiley.com/doi/pdf/10.1002/cctc.201100186 .

17 I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, Energy Environ. Sci. **5**, 9034 (2012).

18 S. Haastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. J. Mortensen, T. Olsen, and K. S. Thygesen, 2D Mater. **5**, 042002 (2018).

19 J. C. Roberts, in *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)* (IEEE, 2007) pp. 61–71.

20 M. d. P. d. S. Maciel, B. S. Meiguins, R. A. d. M. Lourenço, A. S. G. Meiguins, and P. I. A. Godinho, in *2008 12th International Conference Information Visualisation* (2008) pp. 113–119.

21 M. Scherr (2009).

22 J. Sjöbergh and Y. Tanaka, in *2014 18th International Conference on Information Visualisation* (IEEE, 2014) pp. 170–175.

23 Y. Tanaka, J. Sjöbergh, and K. Takahashi, in *2016 20th International Conference Information Visualisation (IV)* (IEEE, 2016) pp. 261–270.

24 https://pandas.pydata.org/.

25 https://scikit-learn.org/.

26 https://reactjs.org/.

27 https://bokeh.pydata.org/en/latest/.

28 T. N. Nguyen, T. T. P. Nhat, K. Takimoto, A. Thakur, S. Nishimura, J. Ohyama, I. Miyazato, L. Takahashi, J. Fujima, and K. Takahashi, ACS Catalysis **10**, 921 (2019).