DEVELOPMENT OF INTEGRATED ACTINIDE CHEMISTRY APPLICATION, AACE, FOR ACCELERATION OF ACTINIDE CHEMISTRY EXPERIMENTS

Masahiko Nakase[1*], Takahiro Nishihara[1], Fauzia Hanum Ikhwan[1], Tomohiro Okamura[1] and Kota Matsui[2]

[1]Tokyo Institute of Technology, Institute of Innovative Research, laboratory for Zero-Carbon Energy, 152-8550, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan
[2]Nagoya University Graduate School of Medicine, Department of Biostatistics, 466-8550, 65 Tsurumai-cho, Showa-ku, Nagoya city, Japan
* Corresponding Author

Abstract. Efficient separation of Minor Actinide (MA) from High-Level Liquid Waste (HLLW) is essential in next generation reprocessing systems. To enable this, finding an efficient separation system is inevitable. We employ a chemoinformatic approach to find suitable diluents, particularly fluorinated diluents, due to some advantages. To efficiently find the candidate molecules we developed Acceleration of Actinide Chemistry Experiment (AACE), which can deploy transfer learning (TL) and human-in-the-loop machine learning (HITL-ML). Our approach utilizes Hansen's solubility parameters derived from molecular structures to predict solubility and extractability, create extraction models for MA surrogate Lanthanide (Ln) and MA. This manuscript outlines the methodology for diluent exploration and the function of AACE.

1. Introduction

Separation of the MA, especially the heat-emitting sources, Am, and Cm, effectively reduces the size of the final disposal repository. The separation of trivalent MA from trivalent Ln is difficult in highly acidic conditions. The solubility of MA extractants (*N*-donor extractants) in hydrocarbon diluent is not high, and third phase can be formed. Some hydrofluorocarbon (HFC) and hydrofluoroolefin (HFO) have no flash point, no toxicity, low ozone depletion potential, and are inherently safe. The applicability of fluorinated compounds to solvent extraction as a diluent has already been proved [1-3]. The chemical properties can be tuned by structural modification. There are many possible candidate diluents, hence a systematic survey is required. The conventional exploration requires strategic, systematic, and iterative experiments. The preparation of many chemicals is expensive, and hot experiments generate radioactive wastes; therefore, such limitations must be overcome to accelerate the attempt. To overcome such hurdles, an integrated ML application, AACE was developed and the concept of how to find a suitable diluent with the aid of ML is described in the Fig. 1. The objectives of Machine Learning (ML) in Actinide separation research were set as follows; (1) minimization of the experimental effort, (2) reduction of wastes, (3) Utilization of reported data. Cold experiments require low-cost and many experiments are possible. The surrogate experiments provide insight for hot experiments, but there is an inconsistency since we want to separate MA from the surrogates. Hot experiments generate some radioactive waste and limit experimental opportunities. Also, various elements exist in the actual HLLW situation, and a high radiation dose generates some active species and some degraded products. Sometimes, there are some insoluble residues. An application was developed to make this possible.

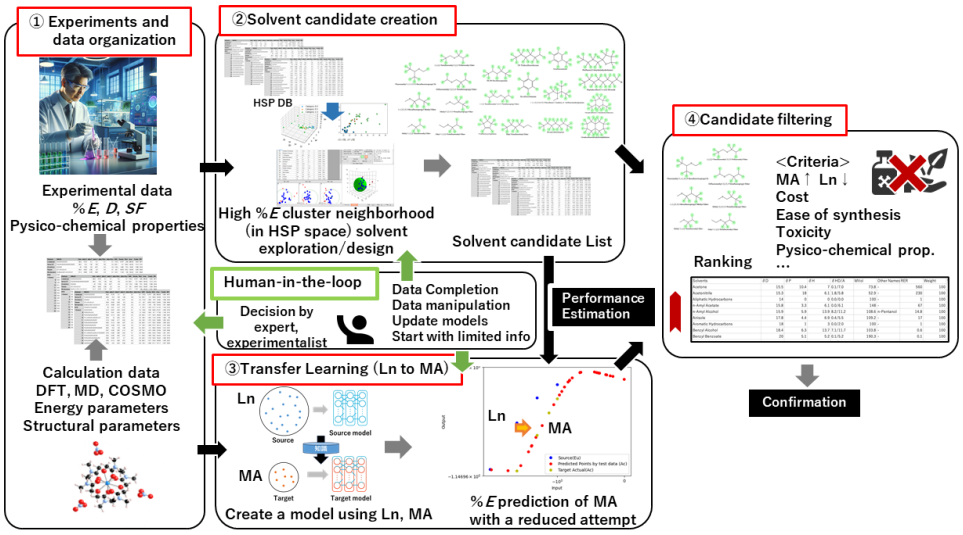


Fig. 1. The concept of how to find a suitable diluent with the aid of ML.

2. AACE program

The AACE program is designed to help experimentalists in actinide chemistry through its multiple functions, such as data manipulation, solubility and extractability prediction, transfer learning to help creating model of target data (Am) from surrogate (Ln), diluent candidate filtration and suggestion (ranking). We adopted an Hansen's Solubility Parameter (HSP)-based approach to find the candidate diluents. The HSP values are the set of three parameters; the dispersion force parameter (δD), the polarity parameter (δP), and the hydrogen bonding parameter (δH) [4]. The parameters used to characterize the solubility of materials, but they are known as a good descriptor to predict some physicochemical properties, and we found this parameter is suitable for finding good diluents.

2.1 Prediction of solubility

A solubility test is essential when implementing solvent extraction. To minimize the experimental effort, filtering out the non-soluble diluent is necessary. As shown in Fig. 2, we first tested some solubility experiments with a typical extractant, tetra-2ethylhexylglycolicacid (T2EHDGA). Obtaining solubility precisely is sometimes challenging and costly; hence, we simplified the test scheme. A 100 mmol/L of T2EHDGA is prepared and tested for respective solubility. The horizontal axis of Fig. 2 indicates the number of carbons in the diluent molecule; blue is soluble, and pink is insoluble.

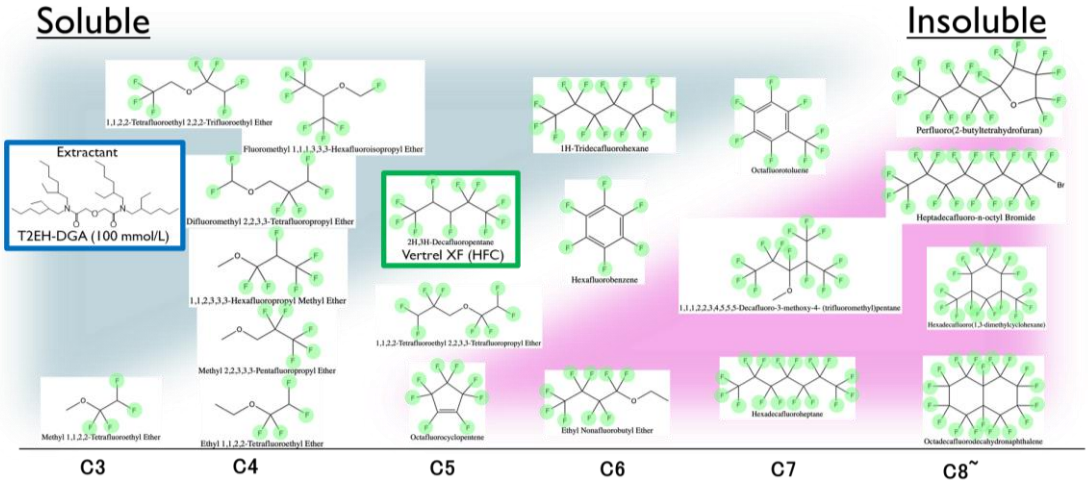


Fig. 2. Preliminary solubility experimental result of T2EHDGA into fluorinated diluents.

There are partially soluble ones, and they are placed in the boundary of the two colours. The result indicates that due to the long and bulky sidechains of T2EHDGA, smaller fluorinated diluents are suitable for dissolution. Then, simple multiple regression was implemented to make this result regression model. Firstly, the values 0, 0.5, and 1 are leveled based on the solubility result, indicating not soluble, partially soluble, and soluble, respectively.

The solubility trends were plotted with the three HSP values corresponding to molecule structures, as shown in Fig. 3. Furthermore, a database of HSP and some correlated estimated values of physicochemical properties can be utilized if needed [5]. The HSP values can be derived by the group contribution method or, more precisely, by DFT calculation. As shown in Fig. 3, the HSP and solubility were correlated by plotting to a 3D graph, suggesting that HSP can filter the solubility of the extractant in the respective diluent. The volume-weighted average can express the HSP values of mixture diluents. Hence, it is possible to find a good mixture using the HSP plot. This is one of the merits of using the HSP values. Sometimes, synthesizing new molecules costs a lot, and differentiating the HSP values is occasionally tricky. The mixture approach can expand the surveillance range of the molecules.

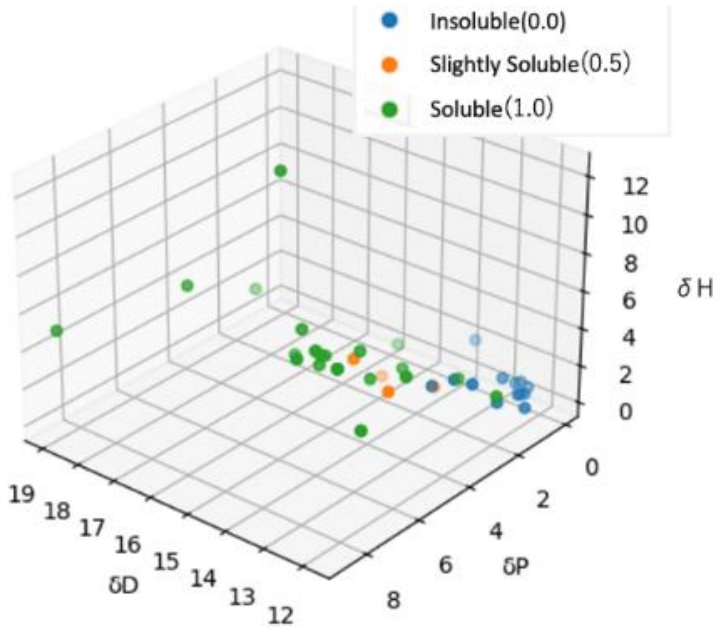


Fig. 3. Clustering of the solubility in HSP space

2.2 Regression model of extractability

Extractability (%*E* or distribution ratio, *D*) is governed by factors, such as extractant structure, concentration of extractant, metal ions and acid, diluent, temperature, contact time, and so on. To simplify the exploration process, we set this study's extractant and nitric acid concentrations to a fixed condition (100 mmol/L of T2EHDGA and 3 mol/L of HNO_3). Firstly, the training data is prepared in CSV format with molecule structures in SMILES (Simplified Molecular-Input Line-Entry System) formula [6]. In addition, each diluent's %*E* and HSP values are calculated by the group contribution method, and some data in a database equipped with HSPiP software and multiple regression parameters are prepared for regression [5]. Then, multiple regression is done to predict the %*E* by Scikit-learn with some models, Ridge, Lasso, Elastic Net, and Gradient-boosted Decision Tree regressions (GBDT) [7]. Fig. 4 illustrates the relation between the %*E* of experiments (x-axis) and prediction values (y-axis). As clearly seen, the GBDT model shows the best performance. However, this model sometimes overfit the result, hence the confirmation by the experiments is needed during the series of the ML processes.

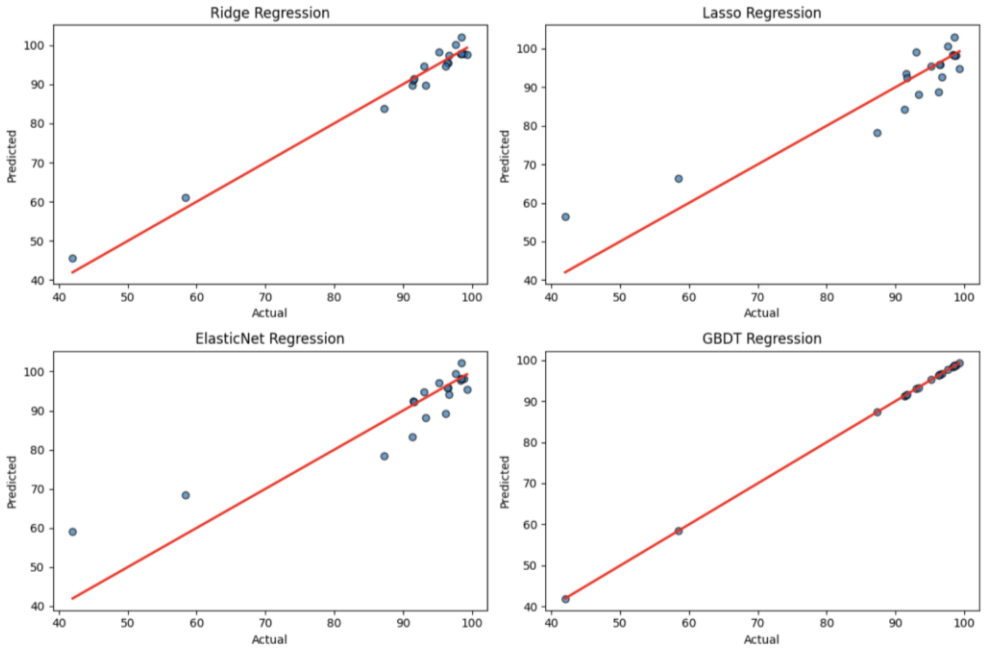


Fig. 4. Regression of %E based on HSP and the derived parameters; (1) Eu and (2) Am.

2.3 Transfer learning (TL) and human-in-the-loop (HITL) concept

TL is a method to enhance the learning efficiency of a target system model by sharing knowledge among systems with similar properties. The initial sparsity of Am data makes it challenging to create an accurate and low-uncertainty %*E* estimation model. To overcome this, we built an Eu model and then constructed an Am model using Eu model since Eu is chemically very similar(surrogate). Therefore, the motivation of TL is to aim at the performance of the prediction model for the Am system with less data by transferring the data and knowledge obtained from the surrogate experiment to the Am system. In the initial stages, expert judgment is necessary for quality assessment, known as HITL. Once the Am model achieves reasonable accuracy, it can identify molecules with higher separation factors (SF) for MA over Ln by the difference of the estimated %*E*s of Eu and Am. Despite the need for more Am extraction experiments to enhance data density, such an acceleration technique by TL and HITL remain beneficial. We used TrAdaboostR2 for transfer learning; Fig. 5 illustrates the effectiveness of using Eu data to improve the Am model. The neural network (NN) trains the data trend, and simultaneous consideration of HSP values improve model validity. The AACE program applies TL using some available schemes, such as TrAdaboostR2 in the ADAPT module [8, 9], and any model that can run on Python can be equipped to expand the AACE functionality. Experiments with Nd and Eu are often performed to simulate Am extraction behaviour in the initial stage before the MA extraction experiments. By modelling the correlation between experimental data (Nd, Eu) from the simulant of Am and actual data of Am, it can be possible to predict the Am behaviours by TL. Also, the AACE suggests the following experimental condition of Am, where the best overall regression conditions are obtained (experimental design). However, issues remain, such as evaluating the predicted extrapolated data derived from the TL. Our approach is not to solve everything with ML but to help experimenters accelerate their research. In each step, the experimentalists can interfere with the ML process (HITL) to make model updates firmly. This HITL concept enables the chemoinformatic approach to start with less experimental data. Fig. 5 illustrates the typical result of TL from %*E* of Eu to that of Am (concentration dependency of T2EHDGA on %*E,* diluent is *n*-dodecane).The Fig 5. (1) shows unsuccessful attempt in TL since the predicted plot in a low DGA concentration region is close to source data (Eu), not target data (Am) even though %*E* of Am of experimental data was successfully interpolated in the higher concentration region of DGA. The Fig.5 (2) shows the successful TL attempt in all the DGA concentration ranges. In the extrapolated region, there is currently no concrete evaluation function; hence, we have to judge by ourselves (HITL). The preliminary error evaluation,

such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2), was summarized in Table 1. If only HSP is used to predict the %*E* of Am, those error evaluation factors suggest the low quality of the prediction. The combination of TL is effective and has the potential for better prediction. However, we must obtain the %*E* of Am in the extrapolated region to check the validity. The model can suggest which experimental condition we should do in the following experiment by combining it with the Gaussian Process (GP) [10]. Quality of the source data (number of plots, errors (fewer outliers)) and model of Eu created by NN. Coverage of source data is crucial in successful TL.

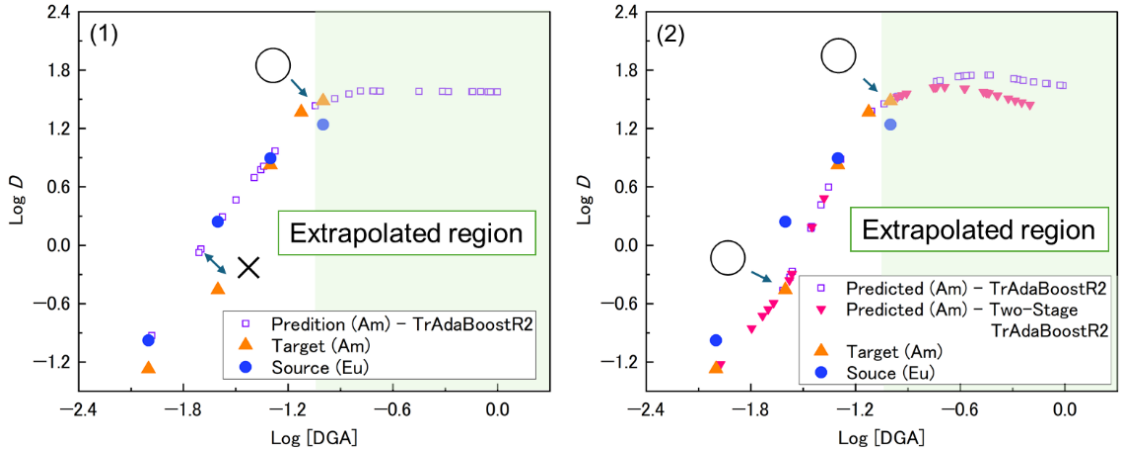


Fig. 5. Example of transition learning deployment from Eu to Am.

Table 1. Accuracy of estimation of D_{Am} from D_{Eu} by Transfer Learning

	Estimation of D_{Am} from D_{Eu}	Estimation of D_{Am} from HSP
MAE	6.629×10^{-3}	1.206
MSE	8.129×10^{-5}	2.221
R^2	9.998×10^{-1}	-1.258

2.4 Suggestion of molecules

The suggestion of diluent molecules (inverse design) for the next experiment is based on the HSP as a feature of the ML model. HSP values is calculated based on the group contribution method, and the plots were divided into five clusters by k-means clustering regarding the extractability, as shown in the left graph of Fig. 6(Red is the higher). Furthermore, the HSP values prepared in the HSPiP software are plotted with the cluster of the %*E* of Eu as shown in the right graph of Fig 6. It is assumed that the diluents near the high %*E* group clusters may show higher %*E*. Diluents within an Euclidean distance of 1 from the center vector of the high extraction cluster were selected as candidates. Filtering based on many factors, such as boiling point and toxicity of these diluents, is possible (HITL approach).

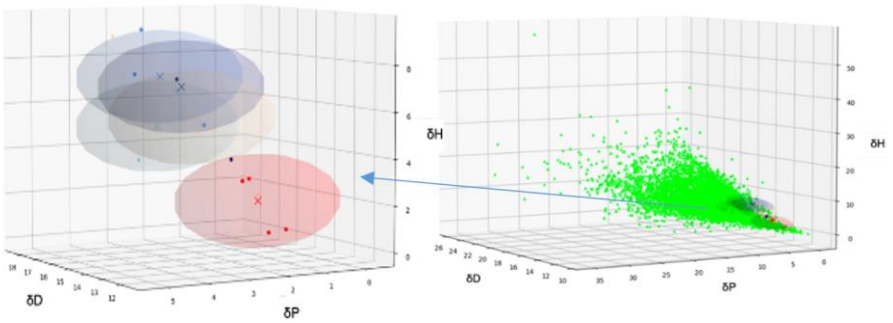


Fig. 6. Filtering of the candidate molecules based on HSP database and clustering

2.5 Reverse engineering for diluent exploration

The suggested molecule and mixture system with Vertrel XF (commercially available fluorinated diluent, 2,3-dihydrodeca- fluoropentane) are shown in Tables 2 and 3, respectively. The scheme predicts the %*E* of Eu, and %*E* of Am from Eu with the aids of TL, and ranks the results according to the difference between the %*E* of Eu and %*E* of Am. Note that the current filtration is based on the fixed concentration of T2EHDGA and detailed solubility is not currently considered.

Table 2. Candidate list of single diluents (partial list)

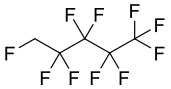
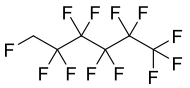
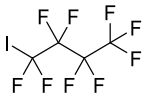
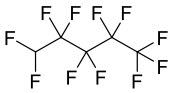
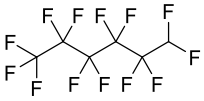
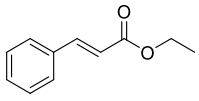
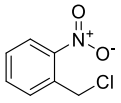
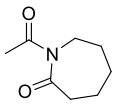
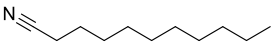
Diluent	SMILES	Chemical structure
1,1,1,2,2,3,3,4,4,5-decafluoropentane	FCC(F)(F)C(F)(F)C(F)(F)C(F)(F)F	
1,1,1,2,2,3,3,4,4,5,5,6-dodecafluorohexane	FCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	
1,1,1,2,2,3,3,4,4-nonafluoro-4-iodobutane	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)I	
1,1,1,2,2,3,3,4,4,5,5-undecafluoropentane	FC(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	
1,1,1,2,2,3,3,4,4,5,5,6,6,7-tetradecafluoroheptane	FCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	
1,1,1,2,2,3,3,4,4,5,5,6,6-tridecafluorohexane	FC(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	

Table 3 Candidate list of Additives to Vertrel XF (partial list)

Mixture diluent	SMILES	Chemical structure	Mix ratio
Ethyl Cinnamate	CCOC(=O)C=CC1=CC=CC=C1	 *The cis and trans conformers are not distinguished	0.24
2-Nitrobenzyl Chloride	C1=CC=C(C(=C1)CCl)[N+](=O)[O-]		0.19
N-Acetyl Caprolactam	CC(=O)N1CCCCCC1=O		0.22
1,1,4-Trifluorobutane	C(CC(F)F)CF		0.39
Undecanenitrile	CCCCCCCCCCC#N		0.38

When the solubility is higher, extractability is also increased, but the selectivity is not drastically changed in the typical extraction condition since the separation factor (*SF*) is defined as the ratio of the distribution ratios (*D*). In this scheme, either %*E* or *D* can be used. Sometime, adjusting the chemical property is hard by structural modification, but if additional chemicals are mixed with diluent, flexibly

in adjustment of the chemical properties can be possible. The mixture diluent system with Vertel XF (structure is illustrated in Fig. 2) was explored based on the fact that the weighted average of the HSP values can represent the mixed diluents. Finally, the experimentalists have to implement the extraction experiments followed by checking the solubility. After that, the regression model is updated by including the new experimental data and continue this flow including ML and inverse design (continuous learning) until the we can find the acceptable extraction system.

2.6 Integrated ML application, AACE

All the functions can work on the Python, but it is useful for researchers who are not familiar with coding if the workflow is not manipulated by a command line. Therefore, AACE integrates ML with a user-friendly interface produced with Streamlit, allowing researchers to work by mouse and click [11]. Fig. 7 shows the snapshot of the diluent exploration by AACE. Many ML functions are installed and users can choose functions depending on the purpose. All the functions are written in Python, but users don't need to modify them. Since Python code runs behind Streamlit, it is easy to add new modules and make changes. As following the process described above, candidates of diluents are suggested. As mentioned above, still the evaluation method of the TL result in extrapolated region is not developed, finally, the validity can be checked by experiments. The progress of ML can be checked during the process, and interference by the human expert is possible since the program progresses step-by-step (implementation of HITL). Current version of AACE even accelerate the attempt to find a good diluent for MA separation, but we are continuously upgrade it followed by the rapid development of ML available by Python.

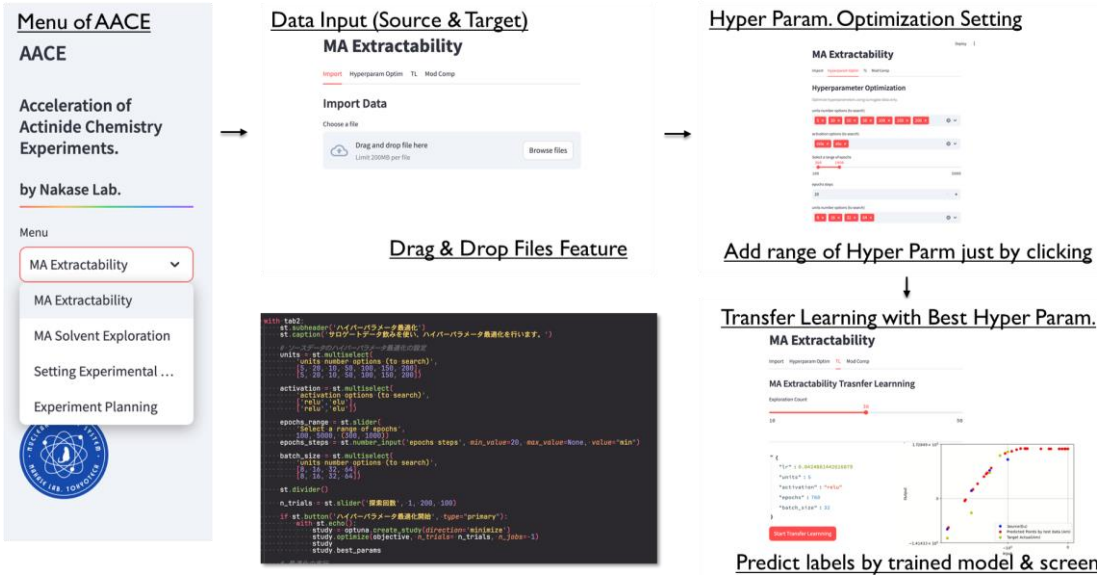


Fig. 7. Snapshot of AACE program.

3. Conclusion

The AACE program was developed to explore the candidate diluents suitable for MA separation. In the AACE program, molecular structures are described in SMILES format, from which HSP values and physicochemical properties are deduced and used for ML. The scheme is based on the distance from clusters with high solubility and high extractability from the molecular structure in the HSP database. In doing so, we have installed the TL to estimate the extraction behaviour of the target Am from the simulant Ln. We have applied prepared scheme to Streamlit so that experimenters can easily implement HITL-ML. Although many improvements are needed in the future, such as a method for evaluating errors in extrapolative estimation in TL and implementation of a more sophisticated solubility determination model, along with further functionality additions, the importance of such efforts has been demonstrated.

Acknowledgements

This study is supported by the MEXT Innovative Nuclear Research and Development Program Grant Number JPMXD0221459189. Actinide extraction experiments were done at the International Research Center for Nuclear Material Science, Institute of Material Research, Tohoku University (Proposal number; 202211-IRKAC-0037, 202212-IRKAC-0413, 202112-IRKAC-0026, 202012-IRKAC-0019, and 20F0023)

References

[1] T. Yamamura, et al., J. Nucl. Sci. Technol., 47, 515-520 (2010)
[2] C. Tabata, et al., Separ. Sci. Technol., 57, 7, 1097-1110 (2021)
[3] K. Sirasaki, et al., RSC Adv., 12, 41, 26922–26933 (2022)
[4] C. M. Hansen, J. Paint Technol., 39, 3, 104-117 (1967)
[5] C. M. Hansen, et al., HSPiP Software (Version 5.4.03), 2021, https://www.hansen-solubility.com/
[6] Weininger, D.. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comp Sci, 28(1), 31-36, 1988, https://doi.org/10.1021/ci00057a005
[7] F. Pedregosa et al., J. Mach. Learn. Res., 12, 2825-2830, 2011
[8] K. Hoffmann and J. Sicking, Adapt: A library for transfer learning (Release 0.4.1) (2021), https://github.com/adapt-python/adapt
[9] Dai, W. et al., Boosting for Transfer Learning, in Proceedings of the 24th International Conference on Machine Learning, ICML '07, 193-200., 2007.
[10] C. E. Rasmussen and K. I. W. Christopher, Gaussian Processes for Machine Learning. Cambridge, MA: The MIT Press, 2006
[11] Streamlit Inc., Streamlit: The fastest way to build data apps, (2023), https://streamlit.io/