Machine learning approach for screening alloy surfaces for stability in catalytic reaction conditions

Gloria A Sulley 1, Jihun Hamm 2 and Matthew M Montemore 1,∗

1 Department of Chemical and Biomolecular Engineering, Tulane University, New Orleans, LA 70118, United States of America
2 Department of Computer Science, Tulane University, New Orleans, LA 70118, United States of America
∗ Author to whom any correspondence should be addressed.
E-mail: mmontemore@tulane.edu

Keywords: alloy, surfaces, stability, machine learning, catalysts

Supplementary material for this article is available online

Abstract

A catalytic surface should be stable under reaction conditions to be effective. However, it takes significant effort to screen many surfaces for their stability, as this requires intensive quantum chemical calculations. To more efficiently estimate stability, we provide a general and data-efficient machine learning (ML) approach to accurately and efficiently predict the surface energies of metal alloy surfaces. Our ML approach introduces an element-centered fingerprint (ECFP) which was used as a vector representation for fitting models for predicting surface formation energies. The ECFP is significantly more accurate than several existing feature sets when applied to dilute alloy surfaces and is competitive with existing feature sets when applied to bulk alloy surfaces or gas-phase molecules. Models using the ECFP as input can be quite general, as we created models with good accuracy over a broad set of bimetallic surfaces including most d-block metals, even with relatively small datasets. For example, using the ECFP, we developed a kernel ridge regression ML model which is able to predict the surface energies of alloys of diverse metal combinations with a mean absolute error of 0.017 eV atom$^{-1}$. Combining this model with an existing model for predicting adsorption energies, we estimated segregation trends of 596 single-atom alloys (SAAs) with and without CO adsorbed on these surfaces. As a simple test of the approach, we identify specific cases where CO does not induce segregation in these SAAs.

1. Introduction

Stability is a crucial performance metric for catalysts, especially for alloys. Therefore, when designing new alloy catalysts, it is important to consider stability under reaction conditions. Accordingly, surface stability has been leveraged in screening catalysts for various important applications [1] such as the electrochemical reduction of CO$_2$ [2], the hydrogen evolution reaction [3], the oxygen reduction [4] and evolution [5] reactions, and CO oxidation [6]. In vacuum conditions, the stability of a metallic alloy surface is related to the formation energy of the alloy surface. Under reaction conditions, the formation energy of the bare surface and the adsorption energies of the relevant intermediates determine the stability. More broadly, surface energies are useful for determining equilibrium nanoparticle shapes and surface segregation [7, 8]. Additionally, surface energies are a fundamental surface property of scientific interest.

Most previous machine-learning (ML) models for heterogeneous catalysis focus on catalytic performance and do not account for stability [9, 10]. Developing models to predict stability could significantly improve efficiency, as surfaces can be simultaneously screened for both catalytic performance and stability. There have been efforts to measure or predict surface stability through experiments [11–13] and quantum chemistry methods such as high-throughput density functional theory (DFT) calculations [14–16] and cluster expansion approaches [17, 18]. Experimental determination of surface energy is a difficult task and often involves extrapolation from the liquid state which poses uncertainties of unknown magnitude [19]. Due to the high computational expense associated with computing these properties using
quantum chemistry methods, the uncertainties from experiments, and the large space of materials available for a single reaction, data driven methods such as ML using DFT calculations as training data could be quite useful for improving the efficiency of these types of predictions.

To predict stability in reaction conditions requires prediction of the energy of the bare surface as well as the adsorption energies of relevant species on the surface. There has been significant progress towards predicting adsorption energies for various reactions using ML methods [20–28]. For example, neural networks [29], extra tree regression [30], convolutional neural networks [31], and kernel ridge regression [32] have been successfully used to study adsorption on various types of materials. Improved feature sets and algorithms for adsorption energies are an active area of research. Nonetheless, less work has been performed towards developing general and efficient models for predicting surface energies of these alloys. Most of these models are tailored for specific kinds of alloys and cannot be used for alloys of many different compositions and structures. Hence, the broad goal of the present work is to develop a more general ML model to predict the formation energies of a broad space of alloy surfaces using ML, as these predictions can be combined with models for predicting adsorption energies to predict stability in reaction conditions.

Past work using ML to predict surface stability [1, 33–35] has resulted in effective models, but generally focuses on a narrow phase space and/or requires a large dataset, which requires significant additional effort for different sets of alloys. For example, recent work [33] used kernel ridge regression in concert with a bond-counting model to predict the surface stability of single atom alloys. While the method is extensible, the dataset only consisted of alloys with up to two dopants. In other work [1], a convolutional neural network was developed to predict cleavage energies of intermetallic alloys. A neural network model [34] has been utilized to predict surface segregation trends in AuPd(111) and AgPd fcc(111) surfaces. Recently, ML has also been used to predict segregation energies of Pt, Ir, Pd, and Rh-based single atom alloys [36]. To create a more broadly useful model, especially one that can predict energies for many compositions and configurations of dilute alloys, it is important to have a dataset that includes many different elements in many different configurations [28]. In this work, we focus on developing a model for surface energies for a broad set of dilute alloy surfaces consisting of nearly all the d-block metals with up to 15 dopants in the top two layers. Our dataset consists of both fcc(111) and hcp(0001) transition metal surfaces.

An important part of developing ML models is to create a suitable feature set, which is a representation of the data set that can be fed into ML algorithms. An effective feature set for our purposes is one that can learn quickly over a dataset with many elements. Many previous feature sets have been developed; here we compare to the Coulomb matrix [37] and its extensions to periodic systems, the Ewald sum matrix (ESM) [38] and sine matrix (SM) [38]. The Coulomb matrix represents the electrostatic interaction between the elements in the system and is well known for predicting properties of organic molecules like atomization energies. These feature sets have shown notable successes, but they are quite low-level descriptions of fundamental physical interactions and hence often require large datasets to reach high accuracy [38]. Improving the data efficiency and generality of ML models for surface formation energies could increase their utility and the overall screening efficiency.

In this current work we develop a ML model using an element-centered fingerprint (ECFP) as the feature set to successfully predict DFT-calculated surface formation energies of a wide variety of alloy combinations in reaction conditions. The ECFP was applied to several other datasets and accuracies were compared with existing feature sets. The ECFP-based models gave good accuracy, even with a small dataset. We combined this method with a previous general model, SurfEP [39], which can be used to predict the adsorption energies of several adsorbates on a variety of alloy surfaces, to predict surface segregation trends in single-atom alloys (SAAs) and hence predict surface stability of these alloy catalysts both with and without adsorbed CO, an important intermediate or poison in many reactions.

2. Methods

2.1. Dataset

The primary dataset used in this work was comprised of alloy surfaces obtained by doping 0–15 atoms into fcc(111) and hcp(0001) transition metal surfaces to give a total of 2543 alloy combinations. The dataset consists of Cu, Ag, Au, Ni, Pt, Pd, Co, Rh, Ir, Fe, Ru, Os, Mn, Re, Cr, Mo, W, V, Ta, Ti, Zr, Hf, and Sc doped into Cu, Ag, Au, Ni, Pt, Pd, Co, Rh, Ir, Ru, Os, Re, Ti, Zr, Hf, and Sc. We utilized three more datasets from previous work to test the validity of ML models. We used a dataset [40] of 5619 dilute stepped Ag alloys obtained by replacing up to eight host Ag atoms with a single transition metal element from the following list: Ti, Zr, Hf, V, Ta, Cr, Mo, W, Mn, Re, Fe, Ru, Co, Rh, Ir, Ni, Pd, Pt, Cu, Au, Zn, Cd. We also used a dataset [41] of 4652 bulk alloy surfaces obtained by binary combinations of 37 selected metals and transition metals in stoichiometric A:B ratios of 0%, 25%, 50%, 75% and 100%. Lastly we tested our method on 7165 organic molecules [37] containing up to 7 atoms of the elements C, O, N, and S, and saturated with hydrogen atoms.
In total, we considered four different datasets with a total of 19 979 systems. Each of these datasets were used separately to train ML models to test the validity of our method.

2.2. DFT calculations
The DFT calculations were performed using the Vienna ab initio Simulation Package [42, 43]. The PW91 exchange-correlation functional [44] was used. The projector augmented wave method [45, 46] was used, with the plane-wave basis set cut off at 396 eV. A $3 \times 3$ surface cell which had four layers was used with a $7 \times 7 \times 1$ k-point mesh. The bottom two layers were fixed. This introduces a physically unrealistic bottom surface, which makes the absolute surface energies less meaningful. However, since our dataset always has the bottom two layers as a pure, bulk-like structure, this contribution is expected to largely cancel out when comparing two surfaces with the same host. Since our goal is to study the stability of dilute alloy surfaces, the relative energy of two surfaces with the same host is the quantity of interest, and we expect this to be relatively reliable. We tested this by running DFT calculations for a Cu$_4$Ag SAA in the top and subsurface layers with four layers (as described above), six layers (with the bottom two layers relaxed) and seven layers (with two symmetric surfaces). The difference in total surface energy between the top layer and subsurface layer per Cu atom was 0.51 eV for seven layers, 0.52 eV for six layers, and 0.61 eV for four layers. Therefore, while this dataset is useful for determining stability, we advise that promising candidates identified during screening using our model should subsequently be tested by running DFT with more layers for those candidates. For ML, surface formation energies were calculated by subtracting out bulk energies of all atoms and normalizing by the number of atoms. For any surfaces with metals that are magnetic in their pure state, spin polarization was used in the calculations. In order to make the data generation tractable, we employed the default spin initialization method for these cases. The force tolerance for relaxation and the electronic structure convergence tolerance were 0.03 eV Å$^{-1}$ and $10^{-7}$ eV respectively. Computational parameters used in the other datasets are available in previous work [37, 40, 41].

2.3. ECFP and ML models
To create a data-efficient and general ML model for formation energies, we developed an ECFP as our feature set. This fingerprint is a concatenation of a set of vectors, one for each element present in any material in the dataset. This fingerprint thus incorporates an important chemical principle: materials and molecules are made up of discrete elements that are modified by their local environment—i.e. by interactions with nearby atoms. The goal is to allow the ML model to efficiently learn how a given element is influenced by its local environment. This contrasts with many feature sets, which treat elements as part of a continuous space.

Therefore, the ECFP is a vector representation of interactions between an element and nearby atoms, where each unique element present in the dataset maps onto a section of the vector (see figure 1). In this work, each element vector encompasses Coulomb interactions between an atom and nearby atoms, inspired by the success of the Coulomb matrix [37]. The Coulomb interaction between an atom and neighboring atoms is calculated, and this is sorted from largest to smallest. This interaction vector is summed over all atoms of the same element. (The standard deviation can also be included to preserve more information, but based on our testing this had little effect on the primary dataset we considered in this work.) Hence, for each element in a given material, we calculate the element vector $V_E$ defined as:

$$V_E = \sum_{i \in E} \text{sort}_1 \left[ \frac{Z_i Z_j(i)}{R_{ij(i)}} \right]$$

where $E$ is the element, $i$ runs over all atoms of that element, $j = j(i)$ runs over the neighbors of $i$, $Z_i$ is the nuclear charge of $i$, and $R_{ij}$ is the distance between $i$ and $j$. The length of $V_E$ is the maximum number of atoms in the local neighborhood, which depends on how the local neighborhood is determined. The neighborhood size can be defined in various ways, such that it is possible to featureize datasets for materials or molecules with quite different structures. In this work, we only used nearest neighbors for surfaces, a major difference between ECFP and the Coulomb matrix, and used all atoms for molecules. For a specific surface, any elements present in the data set but not present in that surface will have $V_E$ as the 0 vector. All $V_E$’s are then concatenated into one vector that is used as the feature set for ML. Hence, this feature set is applicable to datasets with varying numbers of elements, and is likely to scale well to cases with many different elements as compared to the Coulomb matrix which does not have a well-defined placement of elements in the feature set unless other ways of sorting are used [38]. While the Coulomb matrix contains more information than the ECFP, this information may be harder for a ML model to learn on. Specifically, in the Coulomb matrix different elements can occupy the same locations in different matrices, and padding with zeros is required. Therefore, we hypothesize that the Coulomb matrix may give more accurate models for very large datasets than the nearest-neighbor ECFP, especially when few elements are involved or the materials are highly
ordered, but the ECFP will give more accurate models for moderately sized datasets containing many elements. Therefore, the ECFP is expected to give more data efficient and general models, as shown in this work. Our primary goal is to use this vector representation to predict formation energies of a diverse set of alloy surfaces; however, it could also be applied to make predictions of atomization energies or formation energies of many types of molecules and materials.

The ECFP is invariant under rotation and translation, which is important for ML models of materials and molecules. We used Coulomb interactions, but other ways of quantifying the local environment could also be used and may improve the accuracy in specific cases. Our choice of Coulomb interactions was informed by comparisons to other interactions such as the overlap and Hamiltonian matrices, as calculated by a single non-selfconsistent step of the SIESTA DFT code [47]. The Hamiltonian matrix was significantly less accurate when applied to molecules, giving about 53% higher errors. The Coulomb interactions and overlap matrix gave similar errors to each other; however, the Coulomb interactions are computationally convenient since they do not require a call to SIESTA. Hence, Coulomb interactions were chosen for this work, but the ECFP framework can be used with any kind of interaction.

For the development of the ML models, we utilized the kernel ridge regression method as implemented in the scikit learn package [48]. In preliminary experiments, we tested other ML methods such as ordinary least squares, Lasso, ridge regression, kernel ridge regression and support vector regression, and found that kernel ridge regression was the most accurate; hence, kernel ridge regression was used to train all the models. We tested several common kernels while implementing the kernel ridge regression method and found that the Laplacian and Gaussian kernels gave the most promising results; we therefore focus on these two kernels for the purposes of this article. We performed a random 80% training and 20% testing split, performed hyperparameter optimization on the training set using 5-fold cross validation, and then tested the model accuracy on the test set. Specifically, we tuned the regularization parameter and the kernel width. We also tested a splitting scheme that ensured that each bimetallic combination of elements is represented in both the training and testing dataset, but this did not significantly affect the accuracy. To evaluate accuracy, we calculated the mean absolute error (MAE) between the predicted and true values in the test set. We ran the model using five different random seeds and the final MAE was obtained by taking the average of the five.

3. Results and discussion

3.1. ML accuracy
To test the accuracy of the ECFP, we used it to train ML models to predict formation or atomization energies for four different datasets. Specifically, we predicted formation energies for three sets of alloy surfaces: 2543 alloy combinations of most of the d-block metals created by expanding a dataset from previous work [39, 49, 50]; 5619 dilute stepped Ag alloy surfaces taken from previous work [34]; and 4652 bulk alloy surfaces which is a subset of a large database [41] of bimetallic alloy combinations of 37 metals. We also predicted atomization energies of organic molecules using a dataset of 7165 organic molecules obtained from
previous work [37]. These datasets were chosen to test the utility of the ECFP on a diverse set of surfaces and molecules.

Using kernel ridge regression, we created separate ML models for predicting formation energies for each of the surface datasets and atomization energies for the organic molecules. We used the ECFP, as well as feature sets developed in previous work. We utilized three previously developed feature sets: the Coulomb matrix, the ESM and the SM. The Coulomb matrix [37] is a symmetric matrix with off-diagonal elements as the Coulomb interaction between different atoms and diagonal elements as a form of interaction between an atom and itself. The ESM and the SM are extensions of the Coulomb matrix to periodic systems [38].

We find that the ECFP works well on a variety of data sets, especially on dilute alloys (see figure 2). It gives significantly more accurate models than the previous feature sets when applied to dilute alloy surfaces. When using the ECFP to predict surface formation energies using the dataset of 2543 alloys, an MAE of 0.017 eV atom$^{-1}$ was obtained, 60.4% more accurate than the other feature sets. When applied to dilute stepped Ag alloy surfaces, the MAE was even lower, 0.002 eV atom$^{-1}$, and accuracy was again better than previous feature sets. The ECFP gives similar results as previous feature sets on the other data sets: it is slightly less accurate than the existing fingerprints when applied to bulk alloy surfaces, and slightly more accurate than the Coulomb matrix when applied to organic molecules. For the bulk alloy surfaces dataset, the SM gave the most accurate model with an MAE of 0.0084 eV atom$^{-1}$; this is slightly better than the ECFP model which gave an MAE of 0.0097 eV atom$^{-1}$ (see figure 2(c)). When applied to the atomization energy of organic molecules, the ECFP gave an MAE of 8.73 kcal mol$^{-1}$ whereas the Coulomb matrix gave a somewhat higher MAE of 9.59 kcal mol$^{-1}$. We do not show results for ESM and SM for organic molecules because these feature sets are designed for periodic systems. Hence, the ECFP is significantly more accurate than the other feature sets for dilute alloys and gives similar accuracy to the other feature sets in the other cases.

An important property of a ML model is its data efficiency, which is the ability to give good accuracy with relatively small datasets. More data-efficient models can significantly decrease the amount of expensive DFT calculations needed. To test the data efficiency, we calculated learning curves for different feature sets using the 2543 dilute alloy surface dataset. The accuracy of each ML model increases with larger training datasets (figure 3), and the SM does slightly better than the Coulomb matrix and ESM, which agrees with previous work [38]. Interestingly, the ECFP gives relatively accurate models even at small data set sizes. At 500 data points, the ECFP gives an MAE less than 0.06 eV atom$^{-1}$, whereas the MAEs for the other feature sets are
Figure 3. The MAE of using the ECFP, SM, ESM and Coulomb matrix to make predictions at different numbers of training datapoints for the 2543 dilute alloy surfaces dataset.

3.2. Surface segregation
As a simple application of the ECFP model, we predicted surface segregation trends on SAA surfaces with and without adsorbed CO (see figure 4). SAAAs have proven to be effective catalysts for many reactions and understanding their stability in various conditions is a crucial issue. To be an effective catalyst, the single-atom dopant should lie in the surface layer under reaction conditions. To be easily synthesizable in vacuum, the dopant should lie in the surface layer when there are no surface species present. We study the segregation trends of CO on these alloy catalysts because CO is an intermediate in many important reactions and is often used as a probe for surface science and catalysis studies. Moreover, CO can be a major cause of catalyst poisoning in some reactions. Hence, knowing the segregation trends in the presence of CO could inform the choice of catalysts for various reactions.

First, we predicted surface segregation energies for 298 SAA surfaces by comparing the energy of the single dopant atom in the surface and subsurface layers, for a total of 596 predictions. The segregation energy ($\Delta E_{\text{seg}}$) in the absence of any adsorbates was computed by:

$$\Delta E_{\text{seg}} = E_f(\text{sub}) - E_f(\text{surf})$$

where $E_f(\text{sub})$ and $E_f(\text{surf})$ are ML-predicted formation energies of alloys with the single dopant in the subsurface and surface layer, respectively. A positive value of $\Delta E_{\text{seg}}$ indicates that the surface is more stable when the dopant is in the surface layer. In the absence of adsorbates, $\Delta E_{\text{seg}}$ was negative for most cases, suggesting that the dopant is more stable in the subsurface layer. This is consistent with previous work [8, 33]. Further, our results qualitatively reproduce previous DFT findings [8] in all cases but one: RhCu was predicted to have a segregation energy of $+0.03$ eV, when the previous calculations suggest it should be negative; however, the total quantitative error is less than approximately 0.2 eV.

We then predicted the adsorption energies of CO on these 596 surfaces with the dopant in the surface layer and subsurface layer using SurfEP [39], a previously developed model for predicting adsorption energies. SurfEP uses a set of linear submodels to first predict electronic structure parameters of surface atoms, and then predict adsorption energies of a variety of species. While SurfEP can predict adsorption energies of several species, in its current form it cannot predict surface formation energies. We then estimated the segregation energies of these surfaces in the presence of CO as:
Figure 4. (a) and (d) Schematic representations of the different structures interrogated as part of our predictions for surface segregation. (a) Bare surface with dopant located in the top layer. (b) Surface with dopant located in the top layer and adsorbed CO. (c) Bare surface with dopant located in the subsurface. (d) Surface with dopant located in subsurface and adsorbed CO adsorbed. (e) CO-induced segregation energies versus segregation energies of the surfaces without CO.

\[
\Delta E_{\text{seg}}^{\text{CO}} = \Delta E_{\text{seg}} + E_{\text{sub ads}}^{\text{CO}} - E_{\text{surf ads}}^{\text{CO}}
\]

where \(E_{\text{sub ads}}^{\text{CO}}\) and \(E_{\text{surf ads}}^{\text{CO}}\) are the SurfEP-predicted adsorption energies of CO adsorbed on the surface with the dopant in the subsurface and surface layers, respectively. The value of \(\Delta E_{\text{seg}}\) is predicted using our ECFP model, as above. When CO was introduced, we observed that for many Ag, Au and Cu host SAAs, the dopants segregated from the subsurface to the surface. With CO present, our model predicted negative segregation energies for only two cases that were previously calculated to have positive segregation energies, PdAu and RhAu \[8\]. The errors in the predicted CO adsorption energies are the primary drivers of these overall errors, rather than the errors in the surface formation energy predictions.

In this previous work \[8\], DFT calculations found that all 15 of the Au, Ag, and Cu-based SAAs that were tested showed positive surface segregation energies in the presence of CO. This raises the question of whether SAAs with Cu, Ag, and Au always result in stable surface dopants in the presence of CO. In fact, our model predicted several cases where the dopant remained in the subsurface when CO was introduced to the surface. To confirm this, we performed DFT calculations for Sc doped into an Au surface and found that the Sc dopant indeed preferred the subsurface even in the presence of CO, in agreement with our model predictions. Hence, our model could be used to quickly identify SAAs that will not be stable in the presence of a given adsorbate, which can increase the efficiency of screening.

4. Conclusion

We developed a new featurization scheme, the ECFP, in our pursuit to develop a ML model to predict formation energies of alloy surfaces. We trained ML models for surface formation energies by screening multiple matrix-based feature sets and fitting to multiple datasets. We found that the ECFP is able to accurately predict formation energies of a broad swath of alloy surfaces consisting of many different metals. Combining this model with an existing model for predicting adsorption energies we were able to predict surface segregation energies for 298 single atom alloys both with and without CO. Our model predictions showed cases where CO was not able to bring the dopant from the subsurface onto the surface.

Our model is data efficient and can be applied to a variety of datasets. The model can be applied in various fields for making predictions of formation energies where stability is important. Because the model we used for predicting adsorption energies also can make predictions for other species, it is now possible to estimate the stability of a large number (>1 million) of alloy surfaces in the presence of several different adsorbed species. This allows efficient predictions of stability of many possible alloy surfaces without requiring any further data generation or fitting. We share an implementation of our model using the primary dataset discussed in this work (bitbucket.org/mmmontemore/surfep). Hence, fast screening of alloy surfaces for both stability and catalytic performance for many reactions is now possible.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).
Acknowledgments

This work was partially supported by the US National Science Foundation under Grant No. OIA-1946231 and the Louisiana Board of Regents for the Louisiana Materials Design Alliance (LAMDA). Acknowledgment is made to the donors of The American Chemical Society Petroleum Research Fund for partial support of this research (61331-DN15). This research was supported in part using computing resources provided by Technology Services at Tulane University, New Orleans, LA.

ORCID iD

Matthew M Montemore https://orcid.org/0000-0002-4157-1745

References

[1] Palizhati A, Zhong W, Tran K, Back S and Ulissi Z W 2019 Toward predicting intermetallics surface properties with high-throughput DFT and convolutional neural networks J. Chem. Inf. Model. 59 4742–9
[2] Weitzen S E et al 2021 Evaluating the stability and activity of dilute Cu-based alloys for electrochemical CO2 reduction J. Chem. Phys. 155 114702
[3] Greeley J, Jaramillo T F, Ronde J, Chorkendorff I and Nørskov J K 2006 Computational high-throughput screening of electrocatalytic materials for hydrogen evolution Nat. Mater. 5 909–13
[4] Lin R, Cai X, Zeng H and Yu Z 2018 Stability of high-performance Pt-based catalysts for oxygen reduction reactions Adv. Mater. 30 1–8
[5] An L, Yang F, Fu C, Cai X, Shen S, Xia G, Li J, Du Y, Luo L and Zhang J 2022 A functionally stable RuMn electrocatalyst for oxygen evolution reaction in acid Adv. Funct. Mater. 32 1–8
[6] Wang Q et al 2022 Co-promotion of two-type active sites: PtCux single-atom alloy and copper-ceria interface for preferential oxidation of CO Appl. Catal. B Environ. 306 121117
[7] Nilekar A U, Ruban A V and Mavrikakis M 2009 Surface segregation energies in low-index open surfaces of bimetallic transition metal alloys Surf. Sci. 603 91–96
[8] Darby M T, Sykes E C H, Michaelides A and Stamatakis M 2018 Carbon monoxide poisoning resistance and structural stability of single atom alloys Top. Catal. 61 428–38
[9] Kitchin J R 2018 Machine learning in catalysis Nat. Catal. 1 230–2
[10] Schlexer Lamoureux P, Winther K T, Garrido Torres J A, Streibel V, Zhao M, Bajdich M, Abild-Pedersen F and Bligaard T 2019 Machine learning for computational heterogeneous catalysis ChemCatChem 11 3581–601
[11] Métis J I and Müller P 2004 Absolute surface energy determination Surf. Sci. 548 13–21
[12] Messmer C and Bilello J C 1981 The surface energy of Si, GaAs, and GaP J. Appl. Phys. 52 4623–9
[13] Tynoczoj J, Calle-Vallejo F, Colic V, Schuhmann W and Bandarenka A S 2015 Evaluation of the electrochemical stability of model Cu-Pt(111) near-surface alloy catalysts Electrochim. Acta 179 469–74
[14] Tran R, Xu Z, Radhakrishnan B, Winston D, Sun W, Persson K A and Ong S P 2016 Surface energies of elemental crystals Sci. Data 3 160080
[15] Farsi L and Deskins N A 2019 First principles analysis of surface dependent segregation in bimetallic alloys Phys. Chem. Chem. Phys. 21 23626–37
[16] Yu Y, Zhang J, Xiao W, Wang J and Wang L 2017 First-principles study of surface segregation in bimetallic Ni3M (M = Mo, Co, Fe) alloys with chemisorbed atomic oxygen Phys. Status Solidi Basic Res. 254 1600810
[17] Ekborg-Tanner P and Erhart P 2021 Hydrogen-driven surface segregation in Pd alloys from atomic-scale simulations J. Phys. Chem. C 125 17248–60
[18] Vignola E, Steinmann S N, Le Mapihan K, Vandegehuchte M D, Curulla D and Sautet P 2018 Acetylene adsorption on Pd-Ag alloys: evidence for limited island formation and strong reverse segregation from Monte Carlo simulations J. Phys. Chem. C 122 15456–63
[19] Vitos L, Ruban A V, Skriver H L and Kollár J 1998 The surface energy of metals Surf. Sci. 411 186–202
[20] Arrnith N, Lin Z and Chen J G 2020 Predicting the activity and selectivity of bimetallic metal catalysts for ethanol reforming using machine learning ACS Catal. 10 9438–44
[21] Chen C, Ye W, Zuo Y, Zheng C and Ong S P 2019 Graph networks as a universal machine learning framework for molecules and crystals Chem. Mater. 31 3564–72
[22] Gao W, Chen Y, Li B, Liu S-P, Liu X and Jiang Q 2020 Determining the adsorption energies of small molecules with the intrinsic properties of adsorbates and substrates Nat. Commun. 11 1196
[23] Roy D, Mandal S C and Pathak B 2021 Machine learning-driven high-throughput screening of alloy-based catalysts for selective CO2 hydrogenation to methanol ACS Appl. Mater. Interfaces 13 56151–63
[24] Chowdhury A J, Yang W, Walker E, Mamun O, Heyden A and Terejanu G A 2018 Prediction of adsorption energies for chemical species on metal catalyst surfaces using machine learning J. Phys. Chem. C 122 28142–50
[25] Praveen C S and Comas-Vives A 2020 Design of an accurate machine learning algorithm to predict the binding energies of several adsorbates on multiple sites of metal surfaces ChemCatChem 12 4611–7
[26] Chowdhury A J, Yang W, Heyden A and Terejanu G A 2021 Comparative study on the machine learning-based prediction of adsorption energies for ring and chain species on metal catalyst surfaces J. Phys. Chem. C 125 17742–8
[27] Guan Y, Chaffart D, Liu G, Tan Z, Zhang D, Wang Y, Li J and Ricardez-Sandoval I 2022 Machine learning in solid heterogeneous catalysis: recent developments, challenges and perspectives Chem. Eng. Sci. 248 117224
[28] Sulley G A and Montemore M 2022 Recent progress towards a universal machine learning model for reaction energetics in heterogeneous catalysis Curr. Opin. Chem. Eng. 36 100821
[29] Chowdhury A J, Yang W, Abdelfatah K E, Zare M, Heyden A and Terejanu G A 2020 A multiple filter based neural network approach to the extrapolation of adsorption energies on metal surfaces for catalysis applications J. Chem. Theory Comput. 16 1105–14
[30] Toyao T, Suzuki K, Kikuchi S, Takakusagi S, Shimizu K and Takigawa I 2018 Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys J. Phys. Chem. C 122 8315–26
[31] Back S, Tran K and Ulissi Z W 2019 Toward a design of active oxygen evolution catalysts: insights from automated density functional theory calculations and machine learning ACS Catal. 9 7651–9
[32] Jäger M O J, Morooka E V, Federici Canova F, Himanen L and Foster A S 2018 Machine learning hydrogen adsorption on nanoclusters through structural descriptors npj Comput. Mater. 4 37
[33] Rao K K, Do Q K, Pham K, Maiti D and Grabow L C 2020 Extendable machine learning model for the stability of single atom alloys Top. Catal. 63 728–41
[34] Boes J R and Kitchin J R 2017 Modeling segregation on AuPd(111) surfaces with density functional theory and monte carlo simulations J. Phys. Chem. C 121 3479–87
[35] Prabhu A M and Choksi T S 2022 Data-driven methods to predict the stability metrics of catalytic nanoparticles Curr. Opin. Chem. Eng. 36 100797
[36] Salem M, Cowan M J and Mpourmpakis G 2022 Predicting segregation energy in single atom alloys using physics and machine learning ACS Omega 7 4471–81
[37] Rupp M, Tkatchenko A, Müller K R and Von Lilienfeld O A 2012 Fast and accurate modeling of molecular atomization energies with machine learning Phys. Rev. Lett. 108 058301
[38] Faber F, Lindmaa A, Von Lilienfeld O A and Armiento R 2015 Crystal structure representations for machine learning models of formation energies Int. J. Quantum Chem. 115 1094–101
[39] Montemore M M, Nwaokorie C F and Kayode G O 2020 General screening of surface alloys for catalysis Catal. Sci. Technol. 10 4467–76
[40] Hoyt R A, Montemore M M, Fampiou I, Chen W, Tiritakis G and Kaxiras E 2019 Machine learning prediction of H adsorption energies on Ag alloys J. Chem. Inf. Model. 59 1357–65
[41] Mamun O, Winther K T, Boes J R and Bligaard T 2019 High-throughput calculations of catalytic properties of bimetallic alloy surfaces Sci. Data 6 76
[42] Kresse G and Furthmüller J 1996 Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set Comput. Mater. Sci. 6 15–50
[43] Kresse G 1995 Ab initio molecular dynamics for liquid metals J. Non-Cryst. Solids 192–193 222–9
[44] Perdew J P, Chevary J A, Vosko S H, Jackson K A, Pederson M R, Singh D J and Fiolhais C 1992 Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation Phys. Rev. B 46 6671–87
[45] Blöchl P E 1994 Projector augmented-wave method Phys. Rev. B 50 17953–79
[46] Kresse G and Joubert D 1999 From ultrasoft pseudopotentials to the projector augmented-wave method Phys. Rev. B 59 1758–75
[47] Soler J M, Artacho E, Gale J D, García A, Junquera J, Ordejón P and Sánchez-Portal D 2002 The SIESTA method for ab initio order- N materials simulation J. Phys. Condens. Matter 14 2745–79
[48] Pedregosa F et al 2012 Scikit-learn: machine learning in python J. Mach. Learn. Res 12 2825–30
[49] Montemore M M and Medlin J W 2014 A unified picture of adsorption on transition metals through different atoms J. Am. Chem. Soc. 136 9272–5
[50] Montemore M M and Medlin J W 2014 Predicting and comparing C–M and O–M bond strengths for adsorption on transition metal surfaces J. Phys. Chem. C 118 2666–72
[51] Hansen K, Montavon G, Biegler F, Fazl S, Rupp M, Scheffler M, Von Lilienfeld O A, Tkatchenko A and Müller K R 2013 Assessment and validation of machine learning methods for predicting molecular atomization energies J. Chem. Theory Comput. 9 3404–19