Catalysis Today xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Catalysis Today

journal homepage: www.elsevier.com/locate/cattod

Feature engineering of machine-learning chemisorption models for catalyst design

Zheng Li, Xianfeng Ma, Hongliang Xin*

Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form 8 April 2016
Accepted 11 April 2016
Available online xxx

Keywords:
Machine learning
Artificial neural network
Feature engineering
Density functional theory
Electrochemical CO_2 reduction

ABSTRACT

We integrate machine-learning algorithms into the descriptor-based design approach for rapid screening of transition-metal catalysts. By engineering numerical representation of surface metal atoms using easily accessible features such as the local electronegativity and the effective coordination number that are dependent on the surroundings of an adsorption site, together with the intrinsic properties of active metal atoms including the electronegativity, ionic potential, and electron affinity, the machine-learning model optimized with ~250 *ab initio* adsorption energies on bimetallic alloys can capture complex, non-linear adsorbate/substrate interactions with the root mean squared errors (RMSE) ~0.12 eV. We applied the model to search for {100}-terminated multimetallic copper (Cu) catalysts for electrochemical CO_2 reduction where the *CO adsorption energy represents an important efficiency metric. Compared with the traditional high-throughput computational and experimental trial-and-error approach, the machine-learning chemisorption models have great potential in accelerating the discovery of interesting catalytic materials. As the complexity of catalyst structures increases, new features will be needed to learn underlying correlations and avoid introducing significant errors on top of the average DFT prediction errors expected with standard semi-local generalized gradient approximation (GGA) functionals.

1. Introduction

Transition metals are versatile catalysts in directing molecular transformations. For a given metal, its surface reactivity can be tailored by varying geometric characteristics (e.g., coordination, strain), metal ligands, and extrinsic factors (e.g., solvation, support) in the vicinity of active sites [1]. With hierarchical complexities in catalyst design, a priori estimation of chemical reactivity of surface metal atoms is attractive. While all industrial catalysts used today were discovered via costly and time-consuming empirical testing, recent developments of density functional theory (DFT) have led to an unprecedented atomic-scale understanding of the complex processes occurring on catalyst surfaces and an identification of potentially improved catalysts from first-principles [2]. However, the immense phase space of catalytic materials spanned by structural and compositional degrees of freedom precludes thorough screening, even with combinatorial quantum-chemical calculations and high-throughput experiments.

Herein we tackle this problem by developing a machine-learning chemisorption model that captures complex, non-linear adsorbate/substrate interactions through artificial neural networks (ANNs) and thus enables large-scale exploration of catalytic materials space. The proposed design strategy is built upon the kinetic analysis of surface reactions widely used to pinpoint underlying mechanisms and identify optimal adsorption properties of catalysts [3]. Within this design framework, adsorption energies of simple molecular fragments (e.g., *H, *CO, *OH, and *N) are reactivity descriptors extracted from the interrelated reaction network. Such simplification can be established because the chemisorption strength of a simple adsorbate on a given type of catalyst surfaces is linearly correlated with the adsorption energies of similar reaction intermediates including transition-state species [4,5], thus allowing us to directly build up relationships between reactivity descriptors and catalytic performance. While quantum-chemical density functional theory (DFT) provides a rigorous formalism [6,7] to compute adsorption properties for a collection of hundreds of atoms with a reasonable accuracy, it is limited by the computational cost and system size. With the development of catalysis informatics, adsorbate/substrate interactions can be described using learning-based models without resorting to quantum-chemical calculations on a case-by-case basis [8–12]. Although the electronic structure characteristics, e.g., moments of the *d*-states distribution

* Corresponding author.
E-mail addresses: zhengl@vt.edu (Z. Li), maxf4@vt.edu (X. Ma), hxin@vt.edu (H. Xin).

http://dx.doi.org/10.1016/j.cattod.2016.04.013
0920-5861/

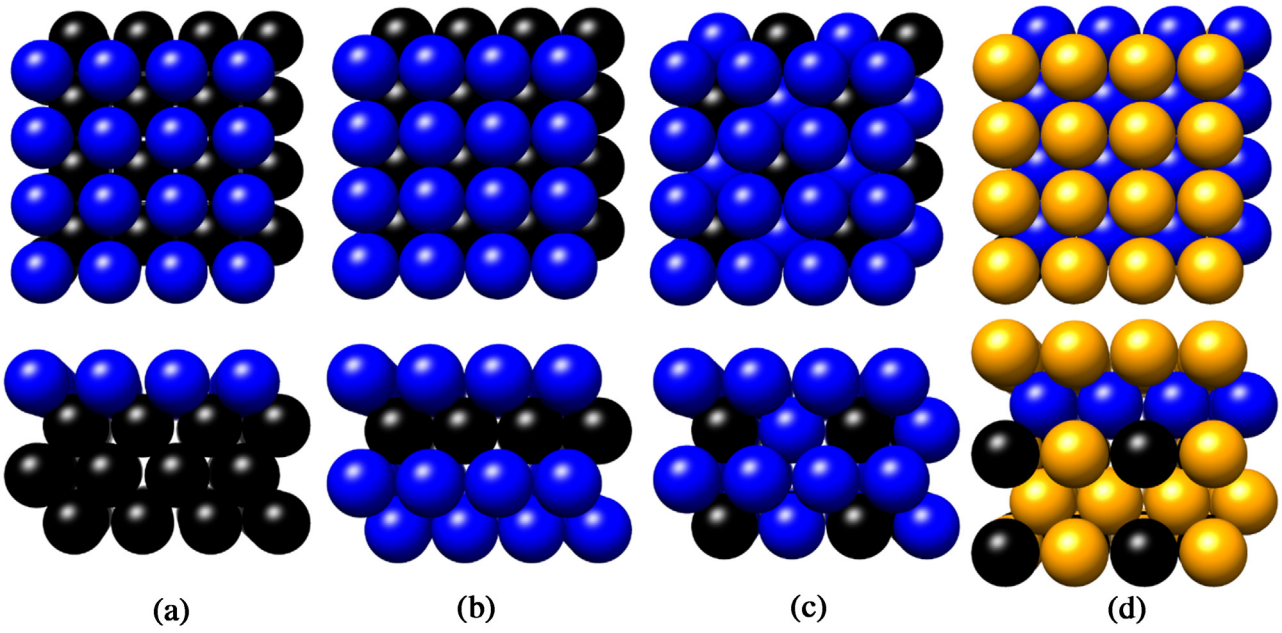

Fig. 1. Structures of {100}-terminated alloy models used in this study. (a) B@A_{ML}, (b) A-B@A_{ML}, (c) A_3B@A_{ML}, and (d) Cu_3B-A@Cu_{ML}. The first and second rows are top and side views of the structures, respectively. Group VIII and IB metals (Cu, Ag, Au, Ni, Pd, and Pt) were selected for both element A and B in the B@A_{ML} structure. In all other model structures, the *d*-block and post transition metals were considered for B, and A is chosen from the group VIII and IB metals.

projected onto an active metal site [13], are the compelling choice as input features in machine learning, the calculations of localized density of states requires convergence of electronic states at dense *k*-point sampling, which prohibits the on-the-fly prediction of catalytic materials. In this contribution, we go one step further to engineer numerical representation of surface atoms in machine-learning chemisorption models using easily accessible features, thereby allowing a facile prediction of surface reactivity of metal catalysts.

2. Computational details

The density functional theory (DFT) calculations were performed using QUANTUM ESPRESSO [14] with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional [15]. We used CO adsorption on the atop site of {100}-terminated monometal and bimetallic surfaces with $4 \times 2 \times 4$ supercells and 15 Å of vacuum space as model systems (core-shell ternary alloys is modelled using $4 \times 2 \times 5$ supercells). The geometric structures of model systems are shown in Fig. 1. The adsorbates and the two topmost metal layers were fully relaxed until the forces were smaller than 0.05 eV/Å while other metal layers were held fixed in their bulk positions. The Brillouin zones of all surfaces were sampled with $4 \times 8 \times 1$ Monkhorst-Pack *k*-points [16]. The kinetic energy cutoff for plane-wave basis sets was 500 eV. The atomic cores were described by ultrasoft pseudopotentials [17]. The occupation of Kohn-Sham eigenstates was smeared by the Fermi-Dirac function with width of 0.1 eV to aid SCF convergence for metallic systems, and all energies were extrapolated to $T = 0$ K.

The CO adsorption energy (ΔE_{CO}) on a surface was calculated by,

$$\Delta E_{CO} = E_{CO/slab} - E_{slab} - E_{CO} \quad (1)$$

where $E_{CO/slab}$, E_{slab}, and E_{CO} are the electronic energies of adsorbate/slab complex, reference slab, and the gas phase CO. The electronic energy of a gas phase CO was corrected by −0.20 eV due to the limitation of the PBE functional for describing gas phase molecules [18].

We use a feedforward artificial neural network (ANN) algorithm implemented in the open-source PyBrain [19] code for a non-linear mapping between the input features and output target function, i.e., the adsorption energies. The details about the network configuration and input features will be discussed in the following sections.

3. Results and discussion

3.1. General strategy for machine-learning accelerated catalyst design

The schematic of the machine-learning accelerated catalyst design approach is shown in Fig. 2. The workflow starts with the data mining of ab initio adsorption energies from materials databases and standardizes the numerical representation of surface active sites using the geometry and composition of the systems. The datasets are then fed into an artificial neural network to build a connection between adsorption properties of catalytic sites and their numerical fingerprints. Once the network learned the underlying correlations of the system, the model can be used to screen through large materials space and suggest materials that give favorable locations on the activity map obtained from the kinetic analysis. Interesting candidates will be evaluated from first-principles before experimental evaluations and the calculated data will be deposited into the database to further improve the model prediction and provide a feedback mechanism of the design cycle.

3.2. Numerical fingerprints of metal surfaces

Feature engineering in machine learning is the process of designing numerical fingerprints of interested systems based on the domain knowledge. Identifying appropriate input features

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013

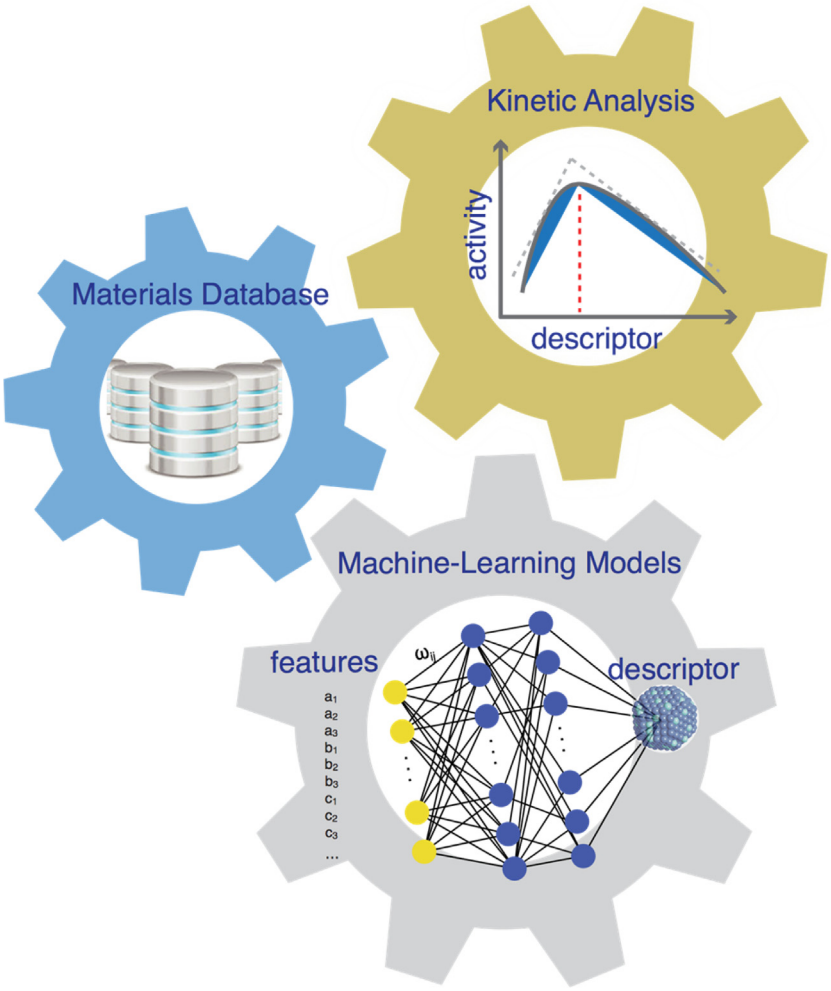

Fig. 2. Schematic representation of the machine-learning accelerated catalyst design approach.

is the most fundamental and challenging step for the application of machine-learning methods. Many different types of features have been proposed to describe the thermodynamic and dynamic properties of materials [20–27]. For instance, the coulomb matrix, based on the Cartesian coordinates and nuclear charges of atoms, has been used as features for describing various thermodynamic properties of molecules and solids [20–22]. The Behler-type symmetry functions are implemented in the high-dimensional neural network potentials for the on-the-fly prediction of interatomic forces in molecular dynamics simulations [23–25]. The partial radial distribution function (PRDF), which considers the distribution of pairwise interactions between two atom types, is used as fingerprints for electronic properties of crystals [26].

Development of physically transparent features that can capture adsorption properties of surface atoms is a key step to the machine-learning accelerated catalyst design. In Fig. 3, we illustrate available features based on our knowledge of the band theory of chemisorption. Among them, the moment characteristics [13] of density of states projected onto the valence *d*-orbitals of a transition-metal atom, e.g., filling, center, width, skewness, and kurtosis, are important factors for understanding reactivity trends of metal surfaces [8,13,28–30]. It has also been shown that the adsorbate-substrate bond length is to a large degree governed by the substrate *sp*-electron density [30]. Due to the difference in the electronegativity of the different metals in the alloys, there is electron transfer between a surface metal atom and its surroundings. To quantify the driving force of electron transfer in metal alloys, we define the local electronegativity of an adsorption site i (χ_i) as the geometric mean of the Pauling electronegativity of metal atoms within the first neighboring shell [8,12],

$$\chi_i = \prod_{j=1}^{1^{st}nn} \chi_j^0 \,^{\frac{1}{N}} \tag{1}$$

where χ_j^0 is the electronegativity of atom j and N is the total number of atoms within the first neighboring shell including the adsorption site i [30]. We consider the local electronegativity as both geometry and electronic structure based feature.

To more efficiently describe the local chemical environment of an adsorption site, we propose the effective coordination number of an adsorption site as another geometry-based features. The effective coordination number of an adsorption site is not a well-defined physical property. A simple bond counting scheme up to 2^{nd} nearest neighbors has been used recently for understanding reactivity trends of pure metal nanoparticles [31]. It has been shown that the effective coordination of under-coordinated surface atoms defined using the accumulated atomic charge density at the site normalized to that of isolated atoms linearly correlates with the core-level shifts [32,33]. Inspired by those studies, we propose the effective

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013

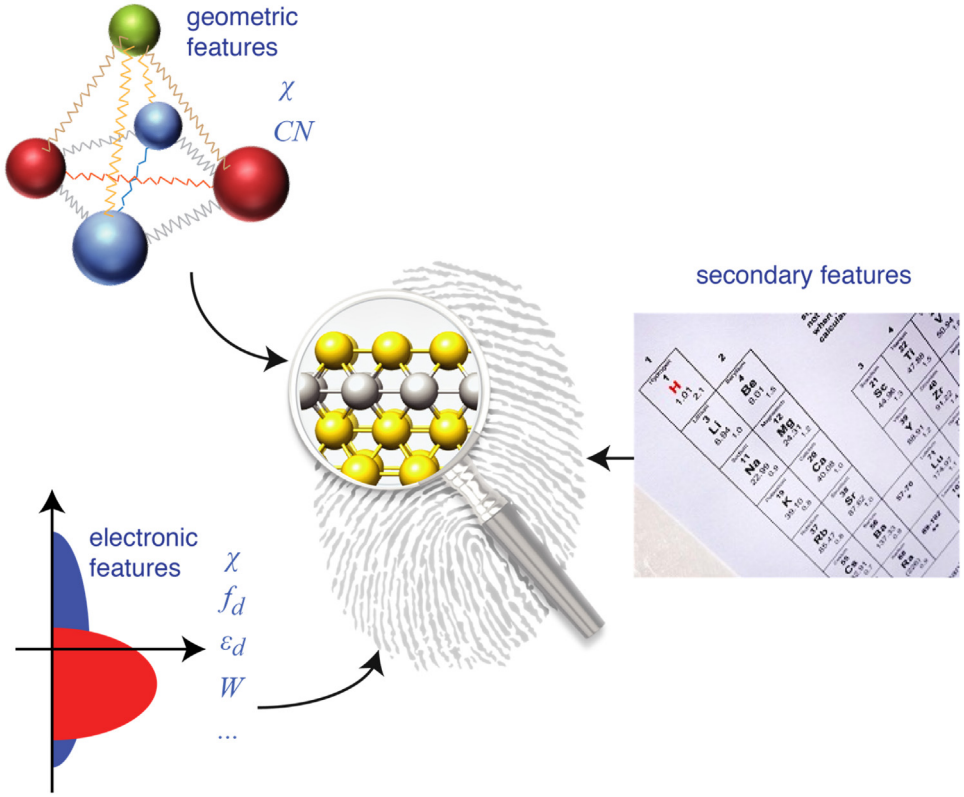

Fig. 3. Feature engineering in machine-learning chemisorption models for metal catalysis. Geometric features: the local electronegativity and the effective coordination numbers, electronic features: the local electronegativity, *d*-band filling, center, width, skewness, and kurtosis, secondary features: the ionic potential, electron affinity, and Pauling electronegativity.

coordination number quantified by the interatomic *d*-*d* coupling matrix elements of an adsorption site with its neighboring atoms up to the second nearest neighbors, defined as,

$$CN_i^d = \sum_{j=1}^{2^{nd}nn} \frac{V_{dd}^{ij}}{V_{dd}^{i,\infty}} \quad (2)$$

where the interatomic *d*-*d* coupling matrix element follows [34], $V_{dd} \propto (r_{d_i}{}^{3/2} r_{d_j}{}^{3/2} / d_{ij}{}^{5})$, r_d is the spatial extent of the metal atom's corresponding *d*-orbital, d_{ij} is the distance between the atom *i* and *j* in the substrate, and $V_{dd}^{i,\infty}$ is the reference interatomic coupling matrix of pure bulk of the host metal with the fully optimized geometry. The *d*-*d* interatomic coupling is chosen because the *d*-band characteristics of a transition metal surface site is largely determining the trend of surface reactivity based on the *d*-band chemisorption theory [35]. In this study, we used the DFT-optimized bond distances for computing the interatomic coupling matrix elements. We note that various classical interatomic potentials, such as the Embedded-Atom Method (EAM) potentials [36], the bond-order potentials [37], the Finnis-Sinclair potentials [38], and the effective-medium potentials [39], can be used to optimize the geometry of large metallic systems very efficiently.

We introduced several secondary features including the ionic potential, electron affinity, and Pauli electronegativity, which are only host-metal dependent and can be fetched from literature or the periodic table. Inclusion of those environment-independent properties is necessary for developing models involving various different metals at adsorption sites.

3.3. Developing machine-learning chemisorption models

3.3.1. Neural network configuration

We set up a feedforward artificial neural network using the open-source PyBrain code [19]. The network consists of a number of layers (input, hidden, and output) and each layer includes a few neurons (nodes) as the processing units. During the training process, each neuron receives the sum of weighted input from neurons of the previous layer and processes the cumulative input through the activation functions. The sigmoid activation function is implemented for the neurons in the hidden layers and a linear function is used for the output layer. The nodes in the previous layer and the subsequent layers are connected by a series of weight parameters, which will be systematically adjusted using the back propagation algorithms in the trainer to minimize the mean squared error between the target function and model hypothesis.

Because the numerical value of different fingerprints could vary in the order of magnitude, feature standardization (scaling of numerical features to have the center at 0 and the standard deviation of 1) is applied routinely to improve the convergence of gradient-based algorithms. 75% of the available dataset is used as the training set and the rest 25% of the dataset is used for testing the model. We aim to prohibit the 'over fitting' of the model, which generally occurs when the network structure becomes excessively complicated (over trained) and fails in making predictions of the unseen data. To determine the optimal network structure with the most suitable number of neurons, we use the built-in Cross Validator module in the PyBrain to do the *k*-fold cross validation for the training set (see Fig. 4a). Using the geometry-based primary features together with secondary features, the network configuration

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013

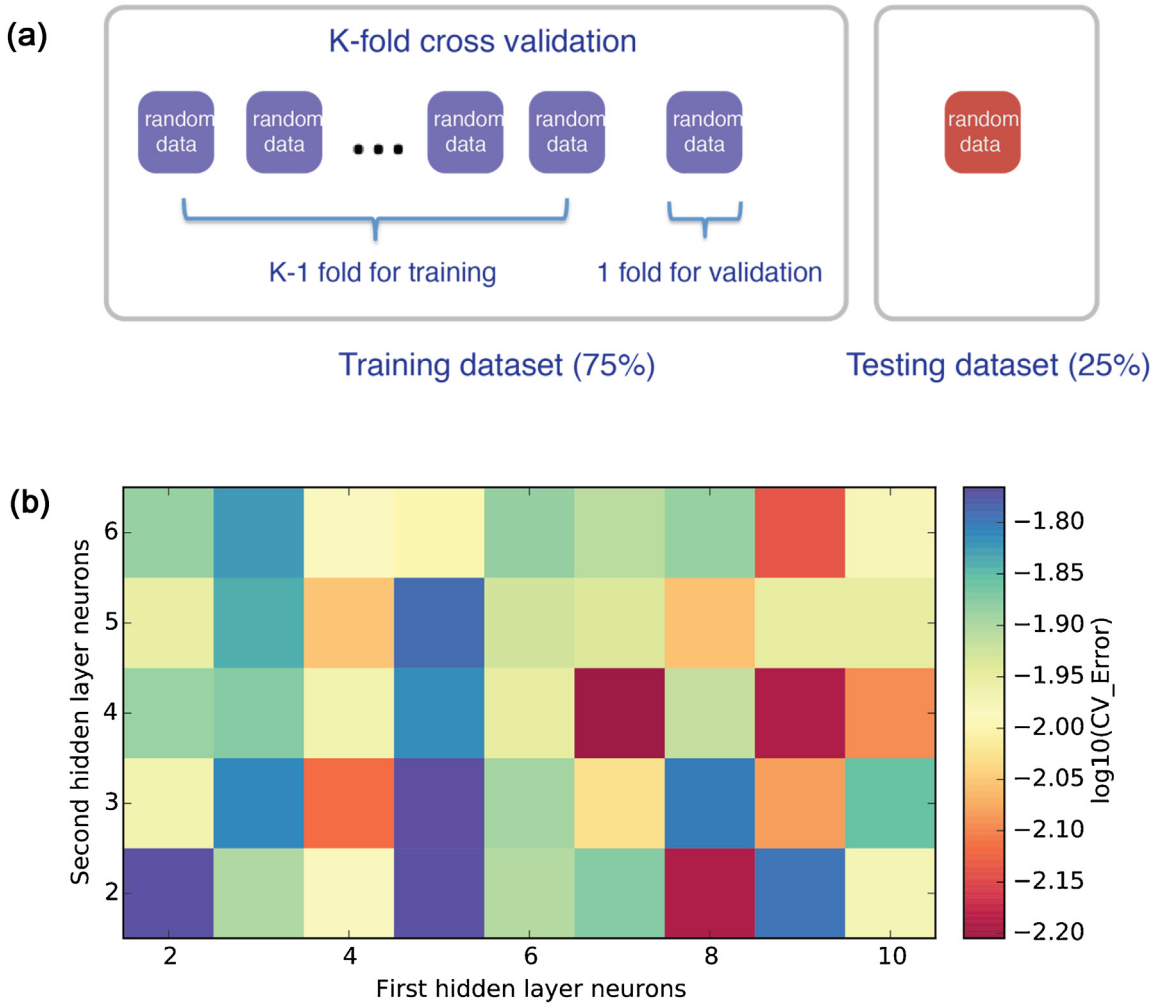

Fig. 4. (a) *K*-fold cross-validation scheme, and (b) the cross-validation errors as a function of the neural network structure using the geometric features and secondary features.

with two hidden layers of 8 and 2 neurons gives the smallest cross validation error (<0.01 eV) (see Fig. 4b).

3.3.2. Comparison of input features

We first compare the neural network performance using the geometric and electronic features together with the secondary features, respectively, as illustrated in Fig. 2. For the geometry-based input features, the network configuration with two hidden layers of 8 and 2 neurons was used, while the two hidden layers of 5 and 2 neurons were identified to give the lowest cross-validation error for the electronic structure based primary features [12]. 16 repeated randomization of data selection and training is used to avoid sampling bias and overly optimistic error estimates. Fig. 5 shows that the geometry-based fingerprints as primary features give similar prediction error (~0.12 eV) compared to that using the electronic features (~0.13 eV). Considering the simplicity of the geometry-based features and easy access to the geometric structure of systems compared to the *d*-band moments, the performance of machine-learning models for describing trends of surface reactivity of bimetallic alloys is excellent. It also suggests that the purely geometry-based local electronegativity and coordination number can be used as indicators of variations in the local chemical environment of an adsorption site in response to strain and ligand engineering.

3.4. Application of machine-learning models for catalyst prediction

We used CO_2 electroreduction on {100}-terminated metal surfaces as a model system to evaluate the performance of the neural network model. Some recent studies on CO/CO_2 electroreduction over Cu(100) suggest that the dimerization of *CO with sequential electron and proton transfers governs onset potentials in C_2 pathways, while a concerted proton-electron transfer to *CO is the critical step in C_1 pathways [18,40]. Using linear scaling relations between adsorption energies of reaction intermediates and the adsorption energy of *CO, we calculate the theoretical limiting potentials for key elementary steps in CO_2 electroreduction along C_1 and C_2 pathways as a function of the CO adsorption energy, shown in Fig. 6(a). Volcano-like reactivity curves for C_1 and C_2 pathways are observed where Cu is close to the top, rationalizing the experimental observation of superior activity of Cu among transition metals for CO_2 electroreduction. Decreasing the binding energy of *CO (less negative) on a Cu surface results in a less negative limiting potential for CO dimerization step. Based on this analysis, the optimal catalysts should have desired *CO adsorption energy in the shaded region, i.e., 0–0.2 eV weaker than that on Cu(100).

To evaluate the applicability of the neural network model with geometric features for catalyst design, we have used the neural-

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013

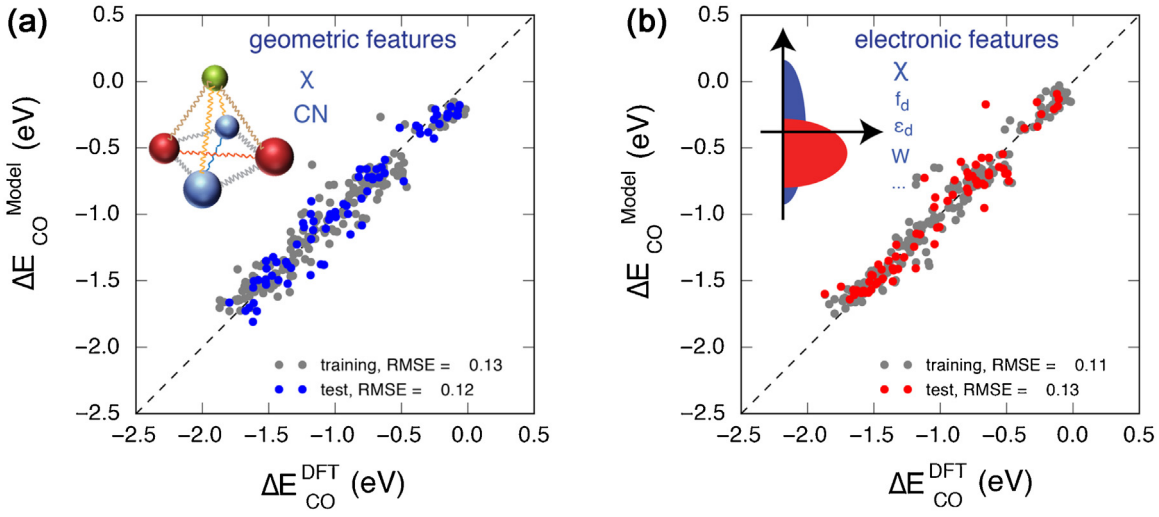

Fig. 5. Comparison of the model performance using the (a) the geometry based, and (b) electronic structured based primary features for CO adsorption energy on metal surfaces. The deviation of the RMSE from 16 random sampling is < 0.02 eV.

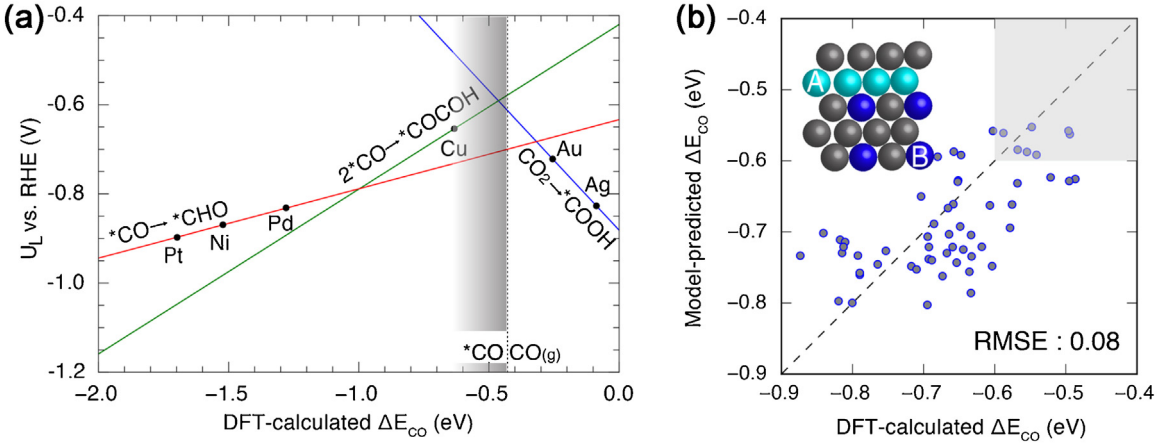

Fig. 6. (a) Predicted limiting potentials for key elementary steps of CO_2 electroreduction to C_1 and C_2 species as a function of CO adsorption energy [12]. A few *d*-block transition metals are overlaid on the map, and (b) the parity plot shows a comparison of the CO adsorption energies on the Cu_3X–Ni@Cu_{ML} and the Cu_3X–Rh@Cu_{ML} alloys (X: transition metal atoms) calculated using the machine-learning model with the geometry-based primary features and the self-consistent DFT. A few alloys including Cu_3Y-Ni@Cu_{ML}, Cu_3Sc-Ni@Cu_{ML}, Cu_3Ti-Rh@Cu_{ML}, Cu_3V-Rh@Cu_{ML} and Cu_3Mo-Rh@Cu_{ML} are identified to have desired CO adsorption energies within the shaded region. The inset (b) shows the geometric structure of the model system. The deviation of the RMSE from 16 random sampling is < 0.02 eV.

network model trained with all available dataset of bimetallics for predicting CO adsorption energy on the 2nd generation core-shell alloys (Cu_3B-A@Cu_{ML}). This type of alloy has a monolayer of Cu metal atoms deposited on a four-layer alloy surfaces (e.g., Cu_3B), and one layer of another guest metal A as a buffer between them. This type of alloy has shown great flexibility in design due to a potential utilization of both strain and ligand effects for tailoring the reactivity of transition metals, and has been extensively studied for many electrochemical reactions, such as oxygen reduction in PEM fuel cells [41,42]. As a simple test, we show model-predicted CO binding energies on a subset of this type of alloys (Cu_3B-Ni@Cu_{ML} and Cu_3B-Rh@Cu_{ML}) as a function of DFT calculations. Fig. 6(b) shows that the neural-network predicted CO adsorption energy agrees reasonably well with self-consistent DFT calculations of CO adsorption (averaged RMSE ~0.1 eV for 16 repeated random sampling in the network training), and several alloys are suggested to have desired CO adsorption energies and thus lower overpotential for CO_2 reduction.

4. Conclusions

In summary, we develop machine-learning chemisorption models using easily accessible, geometric features including the local electronegativity and the effective coordination number of an adsorption site, together with the intrinsic properties of active metal atoms such as the ionic potential, electron affinity, and Pauling electronegativity. Compared with the electronic structure based features, the machine learning model using the new representation of metal surfaces gives favorable prediction error ~0.1 eV for CO adsorption energies. The approach is used for screening {100}-terminated multimetallic alloys for CO_2 electroreduction and shows some promising candidates with lower overpotential. The extension to other types of catalysts, such as realistic multimetallic surfaces and nanoparticles, transition-metal dichalcogenides, and metal oxides, is the subject of ongoing work.

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013

Acknowledgements

H.X. acknowledges the American Chemical Society Petroleum Research Fund (ACS PRF) for partial financial support of this work. The computational work is supported mainly by advanced research computing at Virginia Polytechnic Institute and State University.

References

[1] A. Vojvodic, J.K. Nørskov, Natl. Sci. Rev. 2 (2015) 140.
[2] J.K. Nørskov, F. Abild-Pedersen, F. Studt, T. Bligaard, Proc. Natl. Acad. Sci. 108 (2011) 937.
[3] J.K. Nørskov, T. Bligaard, J. Rossmeisl, C.H. Christensen, Nat. Chem. 1 (2009) 37.
[4] F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T.R. Munter, P.G. Moses, E. Skulason, T. Bligaard, J.K. Norskov, Phys. Rev. Lett. 99 (2007) 016105.
[5] T. Bligaard, J.K. Nørskov, S. Dahl, J. Matthiesen, C.H. Christensen, J. Sehested, J. Catal. 224 (2004) 206.
[6] P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B864.
[7] W. Kohn, L.J. Sham, Phys. Rev. 140 (1965) A1133.
[8] H. Xin, A. Holewinski, S. Linic, ACS Catal. 2 (2012) 12.
[9] A.N. Andriotis, G. Mpourmpakis, S. Broderick, K. Rajan, S. Datta, M. Sunkara, M. Menon, J. Chem. Phys. 140 (2014) 094705.
[10] E.-J. Ras, M.J. Louwerse, M.C. Mittelmeijer-Hazeleger, G. Rothenberg, Phys. Chem. Chem. Phys. 15 (2013) 4436.
[11] M.M. Montemore, J.W. Medlin, J. Phys. Chem. C 118 (2014) 2666.
[12] X. Ma, Z. Li, L.E.K. Achenie, H. Xin, J. Phys. Chem. Lett. 6 (2015) 3528.
[13] H. Xin, A. Vojvodic, J. Voss, J.K. Nørskov, F. Abild-Pedersen, Phys. Rev. B 89 (2014) 115114.
[14] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, A.D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A.P. Seitsonen, A. Smogunov, P. Umari, R.M. Wentzcovitch, J. Phys. Condens. Matter 21 (2009) 395502.
[15] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865.
[16] H.J. Monkhorst, J.D. Pack, Phys. Rev. B 13 (1976) 5188.
[17] D. Vanderbilt, Phys. Rev. B 41 (1990) 7892.
[18] F. Calle-Vallejo, M.T.M. Koper, Angew. Chem. Int. Ed. 52 (2013) 7282.
[19] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, J. Schmidhuber, J. Mach. Learn. Res. 11 (2010) 743.
[20] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O.A. von Lilienfeld, A. Tkatchenko, K.-R. Müller, J. Chem. Theory Comput. 9 (2013) 3404.
[21] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O.A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, J. Phys. Chem. Lett. 6 (2015) 2326.
[22] F. Faber, A. Lindmaa, O.A. von Lilienfeld, R. Armiento, Int. J. Quant. Chem. 115 (2015) 1094.
[23] J. Behler, Int. J. Quant. Chem. 115 (2015) 1032.
[24] J. Behler, J. Phys. Condens. Matter 26 (2014) 183001.
[25] R. Ouyang, Y. Xie, D. Jiang, Nanoscale 7 (2015) 14817.
[26] K.T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K.R. Müller, E.K.U. Gross, Phys. Rev. B 89 (2014) 205118.
[27] B.A. Calfa, J.R. Kitchin, AIChE J. (2016) (n/a).
[28] B. Hammer, J.K. Nørskov, Surf. Sci. 343 (1995) 211.
[29] J.R. Kitchin, J.K. Norskov, M.A. Barteau, J.G. Chen, Phys. Rev. Lett. 93 (2004) 156801.
[30] H. Xin, S. Linic, J. Chem. Phys. 132 (2010) 221101.
[31] F. Calle-Vallejo, J.I. Martínez, J.M. García-Lastra, P. Sautet, D. Loffreda, Angew. Chem. Int. Ed. 53 (2014) 8316.
[32] A. Baraldi, L. Bianchettin, E. Vesselli, S. de Gironcoli, S. Lizzit, L. Petaccia, G. Zampieri, G. Comelli, R. Rosei, New J. Phys. 9 (2007) 143.
[33] L. Bianchettin, A. Baraldi, S. de Gironcoli, E. Vesselli, S. Lizzit, L. Petaccia, G. Comelli, R. Rosei, J. Chem. Phys. 128 (2008) 114706.
[34] W.A. Harrison, Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond, Dover Publications, 1989.
[35] N. İnoğlu, J.R. Kitchin, Mol. Simul. 36 (2010) 633.
[36] M.S. Daw, M.I. Baskes, Phys. Rev. Lett. 50 (1983) 1285.
[37] D.G. Pettifor, I.I. Oleinik, Phys. Rev. B 59 (1999) 8487.
[38] A.P. Sutton, J. Chen, Philos. Mag. Lett. 61 (1990) 139.
[39] K.W. Jacobsen, J.K. Norskov, M.J. Puska, Phys. Rev. B 35 (1987) 7423.
[40] K.J.P. Schouten, Y. Kwon, C.J.M. van der Ham, Z. Qin, M.T.M. Koper, Chem. Sci. 2 (2011) 1902.
[41] K. Gong, D. Su, R.R. Adzic, J. Am. Chem. Soc. 132 (2010) 14364.
[42] H. Yang, Angew. Chem. Int. Ed. 50 (2011) 2674.

Please cite this article in press as: Z. Li, et al., Feature engineering of machine-learning chemisorption models for catalyst design, Catal. Today (2016), http://dx.doi.org/10.1016/j.cattod.2016.04.013