THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

Cite This: J. Phys. Chem. A 2018, 122, 3140−3147

pubs.acs.org/JPCA

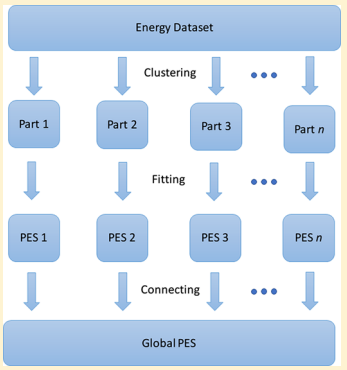
Application of Clustering Algorithms to Partitioning Configuration Space in Fitting Reactive Potential Energy Surfaces

Yafu Guan,[†] Shuo Yang,[†] and Dong H. Zhang*

State Key Laboratory of Molecular Reaction Dynamics and Center for Theoretical Computational Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China

University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

ABSTRACT: A large number of energy points add great difficulty to construct reactive potential energy surfaces (PES). To alleviate this, exemplar-based clustering is applied to partition the configuration space into several smaller parts. The PES of each part can be constructed easily and the global PES is obtained by connecting all of the PESs of small parts. This divide and conquer strategy is first demonstrated in the fitting of PES for OH_3 with Gaussian process regression (GPR) and further applied to construct PESs for CH_5 and O $+CH_4$ with artificial neural networks (NN). The accuracy of PESs is tested by fitting errors and direct comparisons with previous PESs in dynamically important regions. As for OH_3 and CH_5, quantum scattering calculations further validate the global accuracy of newly fitted PESs. The results suggest that partitioning the configuration space by clustering provides a simple and useful method for the construction of PESs for systems that require a large number of energy points.



INTRODUCTION

Born−Oppenheimer potential energy surface (PES) is the most basic quantity to describe a chemical system. Once the PES is constructed, the spectroscopy and reaction properties can be extracted from computer simulations. The PES is defined as a (multidimensional) potential energy function with respect to the molecular configuration coordinates. The construction of PESs involves interpolating[1−6] or fitting[7−10] (multidimensional) potential energy function in a wide range of configuration space.

Among all of the fitting methods, artificial neural networks (NN) have been proven to be a robust and powerful tool to fit accurate PESs for polyatomic molecules in gas phase and for interactions of small molecules with (frozen) surfaces.[11−21] NN is not relying on functional forms derived from physical considerations; instead, it provides simple and flexible functional forms to reproduce ab initio data as accurately as possible.[22] The lack of physical basis leads to very limited extrapolation capabilities. Therefore, it needs to saturate ab initio data in a wide range of configuration space. As a consequence, NN tends to require much more ab initio points to construct satisfactory PES. Hence, substantial effort will be spent on quantum chemistry calculations. A prototypical four atoms PES of OH_3 requires ∼17 000 ab initio energies,[10] and for the prototypical six atoms PES of CH_5, the number is ∼50 000.[23] Large number of ab initio points demand big size of NN to achieve better accuracy. If the NN is too small, subtle features may not be reproduced, which leads to large fitting errors. This underfitting must be avoided.[14] On the other hand, if a bigger NN size is used, it may result in slow convergence of training, making the training process very time-consuming, sometimes even unsolvable.

Recently, another well-established machine learning method, Gaussian processes regression (GPR), was also applied to construct PES for polyatomic molecules.[24−26] Compared to neural networks, GPR shows higher efficiency (i.e., more accurate with fewer points) probably due to the explicit correlations between data points expressed by kernels. However, GPR also has disadvantages. The scalability is poor for large data sets. Gaussian process (GP) inference involves inversion of covariance matrix thus requires $O(n^3)$ time, and predictions are at least $O(n)$ per test point. A large number of training points will lead to slow evaluation of final PES. In addition, GP models commonly adopted use a stationary covariance function (e.g., Matérn covariance function[27]), in which the covariance between any two points only depends on the Euclidean distance and not on their locations directly. The stationary GP model fails to adapt to variable smoothness,[28] which would be a problem in the construction of PESs. For example, in the interaction region, the PES varies more quickly than in the asymptotic region.

Partitioning the configuration space provides a simple but effective divide and conquer solution to all these problems. For artificial neural networks, it can accelerate the training procedure and achieve better accuracy. In the construction of PESs for OH_3 and CH_5, Chen divided the data set into asymptotic parts and interaction parts based on bond lengths.[10,23] The partitioning method Chen adopted has direct chemical meaning, but it is limited to simple chemical reaction systems, in which the related chemical reactions only have one

Received: January 25, 2018
Revised: March 4, 2018
Published: March 7, 2018

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

or two product channels. It is not suitable for a multichannel reaction system like $O(^1D) + CH_4$.[29] As for Gaussian process regression, partitioning can improve the speed of training and evaluation due to the smaller number of data points in each part and also provides a simple but effective nonstationary GP model[30] in the sense that the covariance between any two points of the same Euclidean distance is different in different parts of configuration space.

The partitioning is reminiscent of the clustering in machine learning. Clustering is a type of unsupervised learning in which the goal is to partition a set of points into groups called clusters.[31] If clustering is applied to construct PES, then a quick determination of which cluster a test point belongs to must be provided. Hence, exemplar-based clustering algorithms like *k*-medoids and affinity propagation are possible choices. The task of exemplar-based clustering is to identify a subset of original data points as exemplars to be representatives of clusters. The nonexemplars are then assigned to the nearest exemplar under certain similarity measure. The common similarity measure is the negative squared Euclidean distance, i.e., the nonexemplars are assigned to the nearest exemplar with smallest squared Euclidean distance.

In the present work, the clustering algorithm is applied to partition the configuration space. For each part, NN or GP PES is constructed. The global PES is obtained by connecting all of the PESs through a smooth connection function. This divide and conquer strategy is first demonstrated in the fitting of PES for OH_3 with Gaussian process regression and further applied to construct the PESs for CH_5 and $O + CH_4$ systems with artificial neural networks. The rest of the paper is organized as following. In the Methods section, the details for NN and GP model, the clustering algorithms and the connection scheme are presented. Some properties of the PESs are presented and comparisons are made with earlier surfaces in the Results section. The Summary section contains a brief discussion.

METHODS

Feed Forward Neural Networks. The feed forward NN with two hidden layers connecting the input layer and output layer, denoted as $R - N - M - 1$, is employed. It has R nodes in the input layer, which are equal to the number of dimensions of molecular configuration space, and one node in the output layer corresponding to the potential energy value of the input configuration. The two hidden layers have N and M neurons, respectively. The output of nth neuron in the first hidden layer is

$$a_n^1 = f^1\left(b_n^1 + \sum_{r=1}^{R} w_{n,r}^1 p_r\right) \tag{1}$$

and the output of mth neuron in the second hidden layer is

$$a_m^2 = f^2\left(b_m^2 + \sum_{n=1}^{N} w_{m,n}^2 a_n^1\right) \tag{2}$$

and the final output is given by

$$a = b^3 + \sum_{m=1}^{M} w_m^3 a_m^2 \tag{3}$$

where p_r are the input coordinates, the weight $w_{j,i}^l$ connects the ith neuron of $(l-1)$th layer and the jth neuron of lth layer, and the bias b_j^l determines the threshold of the jth neuron of lth layer, f^1 and f^2 are transfer functions. In this study, $f(x) = x/\sqrt{1+x^2}$ is used as the transfer function[32] which is faster for evaluation than the commonly used hyperbolic tangent function.

The root-mean-square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_{\mathrm{fit}} - E_{\mathrm{target}})^2} \tag{4}$$

is used to measure the performance of NN fitting. The weights and biases are updated through Levenberg–Marquardt algorithm.[33] In addition, the early stopping method[34] was employed to prevent overfitting. To reach the global minimum, 50 times of training with different starting parameters (weights and biases) are performed. NN ensembles are also used to achieve better overall convergence.[23,35−37] In each group of NN training, 10 fittings with the smallest RMSEs are selected, then 5 fittings are selected out of these 10 fittings to form NN ensemble so that averaged PESs can be obtained. In this case, total $C_{10}^5 = 252$ different averaged surfaces are formed and the one with the smallest RMSE is chosen as the final result.

Gaussian Process Regression. A Gaussian process is a collection of random variables $\{\mathbf{f}(\mathbf{x})|\mathbf{x} \in \mathbf{X}\}$, any finite number of which have a joint Gaussian distribution and is completely specified by its mean function $m(\boldsymbol{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$. The covariance function depends on a set of adjustable hyperparameters θ.

For regression problem, suppose we have a training data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ of n noisy realizations of some latent function so that each scalar y_i is obtained by adding Gaussian noise to $f(\mathbf{x})$ at input $\mathbf{x}_i$, i.e., $y_i = f_i + \epsilon_i$, where $\epsilon \sim \mathcal{N}(0, {\sigma_n}^2)$ and $f_i = f(\mathbf{x}_i)$. Let X denote all training inputs, $\mathbf{y}$ denote the corresponding observed targets, X^* denote all n^* test inputs, and $\mathbf{f}^*$ denote the corresponding function values. The prior on the noisy observations $\mathbf{y}$ becomes

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, K(X, X) + {\sigma_n}^2 I) \tag{5}$$

and we can get the joint distribution of $\mathbf{y}$ and $\mathbf{f}^*$ under the prior as

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{f}^*\end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}K(X, X) + {\sigma_n}^2 I & K(X, X^*)\\ K(X^*, X) & K(X^*, X^*)\end{pmatrix}\right) \tag{6}$$

$K(X, X^*) = (k_{i,j^*})$ denotes $n \times n^*$ matrix of the covariances evaluated at all pairs of training and test points and similarly for the other entries $K(X, X)$, $K(X^*, X^*)$, and $K(X^*, X)$.

The prediction step consists of estimating the mean value and the variance for $\mathbf{f}^*$, which are given by the following equations.

$$\mathbf{f}^*|X, \mathbf{y}, X^* \sim \mathcal{N}(\overline{\mathbf{f}^*}, {\sigma_{\mathbf{f}^*}}^2) \tag{7}$$

$$\overline{\mathbf{f}^*} = K(X^*, X)[K(X, X) + {\sigma_n}^2 I]^{-1}\mathbf{y} \tag{8}$$

$${\sigma_{\mathbf{f}^*}}^2 = K(X^*, X^*) - K(X^*, X)[K(X, X) + {\sigma_n}^2 I]^{-1} K(X, X^*) \tag{9}$$

The hyperparameters $\boldsymbol{\theta}$ can be estimated by maximizing the log marginal likelihood given by the following equation:

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

$$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}(K + {\sigma_n}^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log|K + {\sigma_n}^2 I| - \frac{n}{2}\log 2\pi \quad (10)$$

Even though the above GP approach provides an elegant solution to regression problem, it is not suitable for large data sets because training requires $O(n^3)$ time due to the inversion of covariance matrix. The prediction is $O(n)$ for the predictive mean and $O(n^2)$ for the predictive variance per new test case. Thus, large number of training points will lead to slow evaluation of final PES.

Partitioning the Configuration Space. *Exemplar-Based Clustering.* Clustering is an unsupervised learning task of organizing or partitioning data into meaningful groups, which can be applied to partition the configuration space. The *k*-means algorithm is one of the most popular clustering methods. It tries to partition data into *k* clusters in which each data point belongs to the cluster with the nearest mean. The mean serves as the prototype of the corresponding cluster. In the construction of PESs, redundant internal coordinates such as internal bond lengths are commonly used to describe molecular configurations. The mean of such coordinates may not correspond to a physically valid molecular configuration and setting physically invalid configurations as prototypes is not the best option. However, exemplar-based clustering provides an alternative cluster representation by selecting actual data points as the prototype called exemplars.

k-medoids and affinity propagation are the most commonly used exemplar-based clustering methods. *k*-medoids is a variant of *k*-means which is more robust to noises and outliers.[31] Medoid (i.e., exemplar) is the most centrally located point of the cluster, with minimum sum of distances to other points. Partitioning around medoids (PAM) is the classical algorithm for solving the *k*-medoids problem. The basic idea is as follows: Select *k* representative points to form initial clusters (usually done through *k*-means ++ algorithm[38]), and then it searches over all possible swaps between medoids and nonmedoids to see if the sum of medoid to cluster member distances goes down. The algorithm iterates until the medoids do not change.[31] The time complexity of the PAM algorithm is $O(k(N-k)^2T)$, where N is the number of samples and T is the number of iterations until convergence. Thus, PAM is not scalable for large data set. However, some other algorithms have been proposed to improve the efficiency, such as clustering large applications (CLARA)[39] and clustering large applications based upon randomized search (CLARANS).[40] In Statistics and Machine Learning Toolbox of Matlab, a variety of *k*-medoids algorithms are assembled to deal with small to large sized data sets.

Affinity propagation is a clustering method based on the concept of message passing between data points. It could find clusters with much lower error than other methods and does not require the number of clusters to be determined before running the algorithm.[41,42]

Let x_1 through x_n be a set of data points and let s be a function that quantifies the similarity between any two points. The diagonal of s (i.e., $s(i, i)$) referred to as preference is particularly important, which means how likely a particular point is to become an exemplar. When it is set to the same value for all points, it controls how many clusters the algorithm produces.

The algorithm proceeds by alternating two message passing steps, to update two matrices:

(1) The "responsibility" matrix **R** has values $r(i, k)$ that reflects the accumulated evidence for how well-suited x_k is to serve as the exemplar for x_i.

(2) The "availability" matrix **A** contains values $a(i, k)$ that reflects the accumulated evidence for how appropriate it would be for x_i to choose x_k as its exemplar.

Both matrices are initialized to all zeros. The algorithms then performs the following update iteratively:

(a) First, responsibility updates according to the rule:

$$r(i, k) \leftarrow s(i, k) - \max_{k' s.t. k' \neq k} \{a(i, k') + s(i, k')\}$$

(b) Then, availability is updated according to the following relations:

$$a(i, k) \leftarrow \min\left\{0, r(k, k) + \sum_{i' s.t. i' \notin \{i,k\}} \max\{0, r(i', k)\}\right\} \text{ for } i \neq k$$

$$a(k, k) \leftarrow \sum_{i' s.t. i' \neq k} \max\{0, r(i', k)\}$$

The iterations are performed until the exemplar decisions remain unchanged over a number of iterations. The exemplars are extracted from the final matrices as those with $(r(i, i) + a(i, i)) > 0$.

Affinity propagation has a time complexity of the order $O(N^2T)$, where N is the number of samples and T is the number of iterations until convergence. This makes affinity propagation most appropriate for small to medium sized data sets.

Outliers may severely influence the result of clustering, even though the more robust *k*-medoids is used. However, it will be less problematic when partitioning configuration space. To construct PESs that can generate converged dynamical results, it is necessary to saturate data in dynamically relevant regions, which are connected in configuration space. Therefore, the extreme outliers are very unlikely to exist.

In this article, the similarity measure used in both *k*-medoids and affinity propagation is negative squared Euclidean distance, which means that nonexemplars are assigned to exemplar with the smallest squared Euclidean distance.

Connection of PESs. Suppose n exemplars are selected by *k*-medoids or affinity propagation, correspondingly the original training set is divided into n parts. In each part, a NN or GP PES is fitted. The global PES is obtained by connect the PESs of all parts. To make sure the smoothness of PES in the intersection of different parts, we extend every part outward to make it overlap with adjacent parts.

For a nonexemplar, let d_1, d_2, ..., d_n be the distances to each exemplar. By finding the smallest distance d_k, this nonexemplar is then assigned to the kth part. We also assign this nonexemplar to jth part if $d_{j,j\neq k} \leq \rho_0 d_k$, where ρ_0 is a number slightly greater than 1.

During the evaluation process, for a test configuration, let l_1, l_2, ..., l_n be the distances to exemplars. Find the smallest distance l_k, get the energy E_k from kth part, and set its weight to 1. For $l_{j,j\neq k} < \rho_0 l_k$, let $\rho_j = l_{j,j\neq k}/l_k$, get the energy E_j from jth part, and set its weight to

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

$$w(\rho_j) = \frac{1}{2}\left(1 + \cos \pi \frac{\rho_j - 1}{\rho_0 - 1}\right) \quad (11)$$

The final energy is the following weighted mean

$$E = \frac{\sum_{js.t.\rho_j<\rho_0} w(\rho_j) E_j}{\sum_{js.t.\rho_j<\rho_0} w(\rho_j)} \quad (12)$$

RESULTS

GP PES for OH_3. The validity of the proposed strategy is first examined in the fitting of PES for OH_3 with Gaussian process regression.

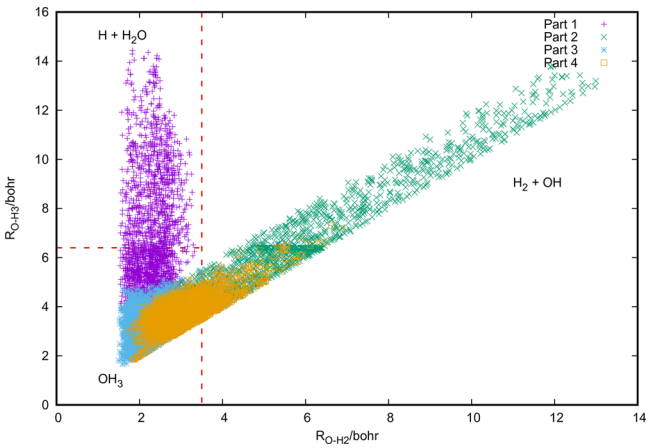


Figure 1. Spatial distribution of geometries for OH_3 as a function of R_{O-H2} and R_{O-H3}.

Table 1. Fitting Results of OH_3

	no. of energy points	fitting error (meV)[a]
part 1	2450	1.52
part 2	1685	1.54
part 3	2485	1.09
part 4	2349	1.37

[a]Root-mean-square error over relevant data set.

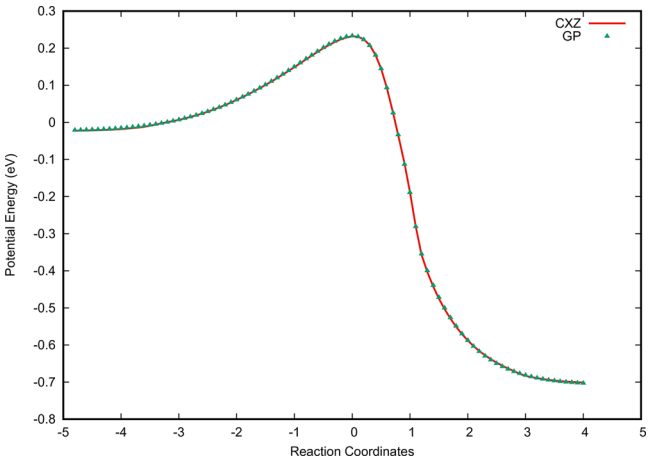


Figure 2. Comparison of the GP PES (green triangles) and the CXZ PES (red line) along the reaction path for $H_2 + OH \leftrightarrow H + H_2O$.

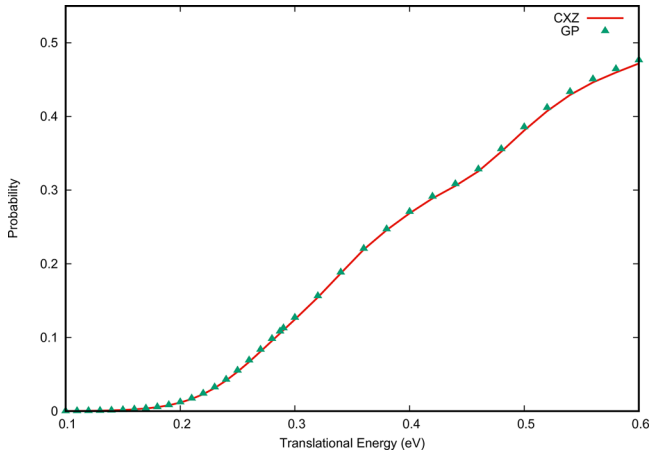


Figure 3. Reaction probabilities for the $H_2 + OH \leftrightarrow H + H_2O$ reaction on the CXZ and GP PESs as a function of collision energy.

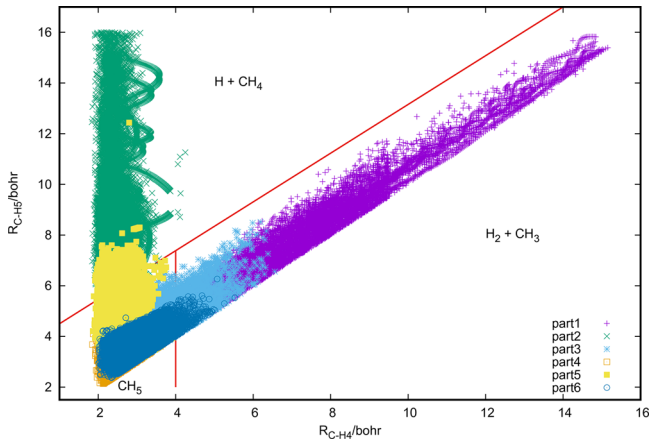


Figure 4. Spatial distribution of geometries as a function of R_{C-H4} and R_{C-H5}.

Table 2. Neural Network Structure Parameters and Fitting Errors for Each Part of CH_5

	NN structure	fitting points	RMSE (meV)
part 1	15−40−40−1	12 541	1.36
part 2	15−40−40−1	20 725	0.94
part 3	15−40−50−1	9979	3.27
part 4	15−55−65−1	8005	8.27
part 5	15−60−65−1	9983	2.02
part 6	15−45−55−1	8439	5.38

The original training data for OH_3 PES were obtained by performing extensive quasi-classical trajectory (QCT) calculations on CXZ PES.[10] Extra points with the emphasis on the low energy $H_2O + H$ and $H_2 + OH$ regions were added. Points with too small Euclidean distances to other existing points were discarded. A total of 8398 points were assembled, which was guaranteed to generate satisfactory PES considering that only 920 energy points were sufficient for GPR to reproduce the dynamic results.[26] The energies were read from CXZ in eV. With all these points used in GPR, the resulting GP PES would be very slow for evaluation. Thus, partitioning the configuration space is indispensable.

The input vector were chosen as the following six bond lengths: R_{OH_1}, R_{OH_2}, R_{OH_3}, $R_{H_1H_2}$, $R_{H_1H_3}$, and $R_{H_2H_3}$. For every

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

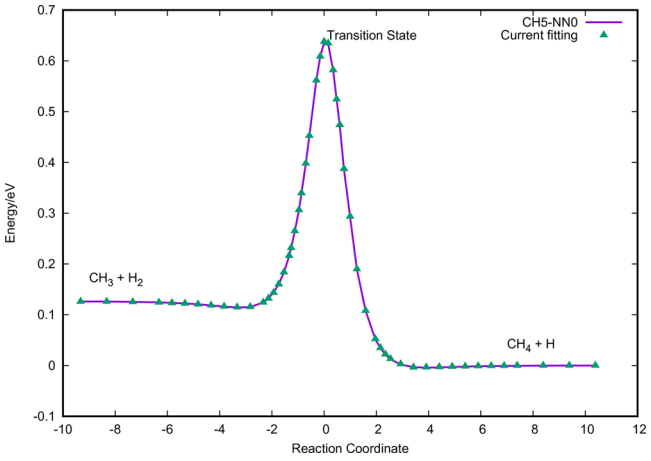


Figure 5. Minimum energy path for reaction H + CH_4 → H_2 + CH_3 for both CH5-NN0 (purple line) and the current fitting (green triangles).

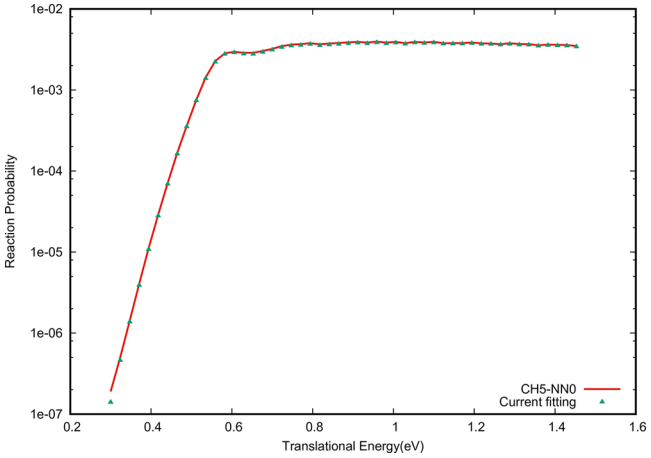


Figure 6. Reaction probabilities for reaction H + CH_4 → H_2 + CH_3 on both CH5-NN0 and the current fitting.

Table 3. Neural Network Structure Parameters and Fitting Errors for each Part of O + CH_4

	NN structure	fitting points	RMSE (meV)
part 1	15−60−80−1	42 231	9.925
part 2	15−50−60−1	34 347	5.064
part 3	15−60−60−1	29 583	7.523
part 4	15−40−50−1	33 231	3.004
part 5	15−60−80−1	20 901	14.37
part 6	15−30−40−1	21 437	6.530
part 7	15−80−80−1	377 54	14.70
part 8	15−60−80−1	18 867	10.09
part 9	15−80−90−1	39 340	17.77
part 10	15−40−50−1	11 433	14.64
part 11	15−60−60−1	33 916	11.89
part 12	15−60−80−1	41 189	7.492
part 13	15−80−90−1	39 562	20.24

data point, three H atoms were permuted in advance to have $R_{OH_1} \le R_{OH_2} \le R_{OH_3}$. The inverse of bond lengths were used in affinity propagation, and four exemplars were found. Correspondingly, S was divided into four parts with ρ_0 = 1.1.

The spatial distribution of geometries for OH_3 is shown in Figure 1. The affinity propagation found one exemplar in the H + H_2O asymptotic region, one exemplar in the H_2 + OH asymptotic region, and two exemplars in the OH_3 interaction region. There were 2450, 1685, 2485, and 2349 points in each part, respectively. With the average number of points being 2241, which is about 1/4 of number of total points, the evaluation of GP PES is expected to be four times faster than that with a total of 8398 points given that the computation time of assigning input configurations to clusters is negligible.

As can be seen in Table 1, the fitting errors are quite small. The terrains of PESs in different parts are different. For example, part 2 is located in the H_2 + OH asymptotic region, where the PES is relatively flat; on the other hand, parts 3 and 4 are located in the OH_3 interaction region, where the PES is more complicated and varies more quickly. As mentioned before, a single stationary GP model may fail to adapt to such drastic change. However, by partitioning the configuration space, a set of stationary GP models is employed to fit the global PES, which can be more efficient.

Figure 2 shows a comparison between GP PES and the CXZ PES along the reaction path for H_2 + OH ↔ H + H_2O. As can be seen, the agreement is quite good, with RMSE along the path being only 1.72 meV. To test the accuracy of GP PES, the calculation of the total reaction probabilities for the H_2 + OH ↔ H + H_2O reaction was performed on the CXZ and GP PES. As shown in Figure 3, the reaction probability for the reaction on these two PESs are very close, thus validating the global accuracy of the GP PES.

Due to the poor scalability, the application of GPR to construct reactive PESs for quantum dynamical simulation is limited to low-dimensional ones.[25] However, by efficient active data selection, the 15 dimensional PES of CH_5 has been reproduced.[26] It can be further anticipated that, in combination with partitioning the configuration space presented here, PES of higher dimensionality can be tackled by GPR.

NN PES for CH_5. Basic training data for CH_5 was obtained from previously fitting of our group.[23] A total of 47 783 energy points were used, and the corresponding NN PES can produce reliable dynamical results, which is denoted by CH5-NN0. In this study, the basic training data set was augmented with extra points generated by running quasi-classical trajectories, and a total of 63 000 geometries were assembled. Affinity propagation is not suitable for this size of data set, instead, *k*-medoids was employed.

The 15 bond lengths were used to describe the geometry of CH_5. The hydrogen atoms were permutated in advance to have $R_{C-H1} \le R_{C-H2} \le R_{C-H3} \le R_{C-H4} \le R_{C-H5}$, and the inverse bond lengths were fed into *k*-medoids clustering. The original training set was divided into six parts with ρ_0 = 1.2. The spatial distribution of geometries is shown in Figure 4. The *k*-medoids found one exemplar in the H + CH_4 asymptotic region, one exemplar in the H_2 + CH_3 region, and four exemplars in the CH_5 interaction region. There were 12 541, 20 725, 9979, 8005, 9983, and 8439 points in each part, respectively. The energies were computed from CH5-NN0 in eV. Each part was fitted by NN, and averaged PES with the smallest RMSE from NN ensemble was chosen as the final result. The fitting results are listed in Table 2.

As can be seen in Table 2, the fitting results can be categorized into three groups: [part 1 (1.36 meV), part 2 (0.94 meV)], [part 3 (3.27 meV), part 5 (2.02 meV)], and [part 4 (8.27 meV), part 6 (5.38 meV)]. Parts 1 and 2 correspond to the H_2 + CH_3 asymptotic region and H + CH_4 asymptotic region respectively where the PES is relatively flat and simple;

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

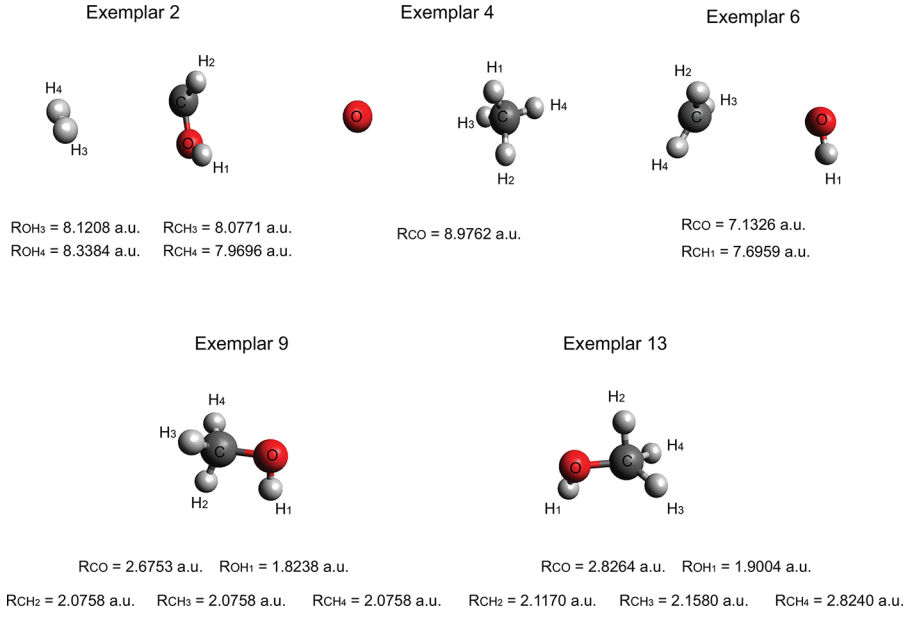


Figure 7. Exemplars for parts 2, 4, 6, 9, and 13 of OCH_4. Exemplar 2, 4, and 6 are in asymptotic regions. Exemplar 9 and 13 are in the interaction region.

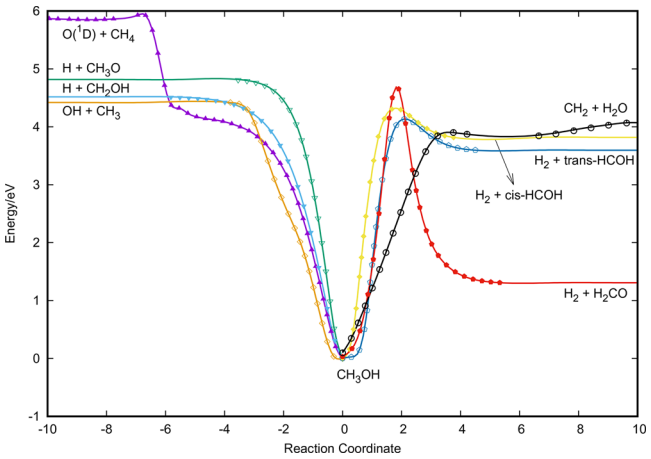


Figure 8. Minimum energy reaction paths of the $O(^1D) + CH_4$ multichannel reaction obtained from the current fitting (symbols) and those obtained from OCH4-PIP (solid lines).

therefore, small errors can be achieved with smaller NN sizes. Due to the complexity of PES in the CH_5 interaction region, parts 4 and 6 have the largest RMSEs, and the corresponding NN sizes are bigger. As for parts 3 and 5, they correspond to the intermediate region between interaction region and asymptotic region, thus having intermediate fitting errors. It can be concluded that more complicated parts need more complicated models to fit.

To analyze the CH_5 PES, Figure 5 shows the minimum energy path through the abstraction transition state for both the CH5-NN0 PES and the current fitting. The agreement is quite good, with the largest difference being 1.4 meV and an RMSE along the path of only 0.58 meV. The calculation of the total reaction probabilities for the $H + CH_5 \leftrightarrow H_2 + CH_3$ reaction was also performed to test the global accuracy of newly fitted PES. In Figure 6, the reaction probability curve is well-reproduced except for a slight difference.

NN PES for O + CH_4. Training data for $O + CH_4$ was obtained from a previous fitting of our group.[29] Roughly 340 000 geometry points were used to fully cover the configuration space of the multichannel reaction system. The energy data was fitted by permutationally invariant polynomials, and the corresponding PES is denoted by OCH4-PIP. The large number of training points can hardly be fitted by a single NN, therefore partitioning the configuration space is also indispensable. Due to the multichannel nature for $O + CH_4$ system, there is no simple partition of the configuration space based on bond lenghts.[10,23] However, the partitioning can be easily done by *k*-medoids.

The 15 bond lengths were used to describe the geometry of $O + CH_4$. The hydrogen atoms were permutated in advance to have $R_{O-H1} \le R_{O-H2} \le R_{O-H3} \le R_{O-H4}$, and the inverse bond lengths were used in *k*-medoids. The energies were computed from OCH4-PIP in eV. The original training set was divided into 13 parts with $\rho_0 = 1.1$. Each part contains 42 231, 34 447, 29 583, 33 231, 20 901, 21 427, 37 754, 18 867, 39 340, 11 433, 33 916, 41 189, and 39 562, respectively. Considering the largest number IS 42 231, a single NN is capable of fitting the PES of each part. Averaged PES with the smallest RMSE from NN ensemble was chosen as the final result.

As shown in Table 3 and Figure 7, quite similar results were obtained for $O + CH_4$. Parts 2, 4, and 6 have the smallest RMSEs: 5.064, 3.004, and 6.530 meV, and the corresponding exemplars are located in asymptotic regions. Parts 9 and 13, on the other hand, have the largest RMSEs: 17.77 and 20.24 meV, and the exemplars are located in interaction regions. Figure 8 shows a comparison between current fitting and OCH4-PIP over reaction paths for various product channels. In general, the agreement is quite satisfactory, with an RMSE along all the paths being 7.3 meV.

SUMMARY

With the steady improvement of quantum chemistry and computational hardware, the treatment of larger and more complex systems is now feasible. The well-established machine learning algorithms such as artificial neural networks and Gaussian process regression also facilitate the construction of

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

PES without making unnecessary approximations. However, large number of energy points adds great difficulty to the construction of PESs. In the present work, this problem was addressed by applying exemplar-based clustering to partition the configuration space into several parts. Exemplar-based clustering tries to find exemplars, members of the input set that are representatives of clusters. The nonexemplars are assigned to the closest exemplar. This provides a quick determination of which cluster a test point belongs to. Each part has relatively smaller number of points, which can be easily handled. The global PES is obtained by connecting the PESs of all parts.

This divide and conquer strategy was first examined in the construction of PES for OH_3 with Gaussian process regression. The first benefit from partitioning the configuration space is the improvement of speed for evaluation due to the smaller number of energy points in each part. Moreover, a set of stationary GP models employed to fit the global PES also provides a simple nonstationary GP model. The accuracy of GP PES was not only tested by fitting errors but also validated through quantum scattering calculations.

The clustering was then further applied to the fitting of PESs for CH_5 and O + CH_4 with artificial neural networks. It is worth mentioning that by partitioning the configuration space in order to make the number of energy points manageable in each part, the PES for O + CH_4 can be fitted, which would otherwise be hardly handled by a single NN with the total number of energy points being ∼340 000. Similar results were obtained that more complex parts need more complex NN models to fit. In general, NN fittings can faithfully reproduce the energies in each part. Quantum scattering calculation was also performed to test the global accuracy of PES for CH_5. In addition, the evaluation speed of NN PESs can also be improved due to the smaller NN size of each part.

To summarize, partitioning the configuration space by exemplar-based clustering provides a general divide and conquer strategy and can extend the construction of reactive PES to systems which requires a large number of points. We expect this strategy will find wide applications in constructing reactive PESs for many systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: zhangdh@dicp.ac.cn.

ORCID

Yafu Guan: 0000-0003-3723-4233

Author Contributions

†Y.G. and S.Y. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Umair Umer for his help checking the manuscript. This work was supported by the National Natural Science Foundation of China (Grant Nos. 21688102, 21590804, and 21433009) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB17010200).

REFERENCES

(1) Xu, C.; Xie, D.; Zhang, D. A Global ab initio Potential Energy Surface for F+H2 → HF+H. *Chin. J. Chem. Phys.* **2006**, *19*, 96−98.

(2) Qiu, M.; Ren, Z.; Che, L.; Dai, D.; Harich, S. A.; Wang, X.; Yang, X.; Xu, C.; Xie, D.; Gustafsson, M.; et al. Observation of Feshbach Resonances in the F + H2 → HF + H Reaction. *Science* **2006**, *311*, 1440−1443.

(3) Ischtwan, J.; Collins, M. A. Molecular potential energy surfaces by interpolation. *J. Chem. Phys.* **1994**, *100*, 8080−8088.

(4) Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. Molecular potential energy surfaces by interpolation in Cartesian coordinates. *J. Chem. Phys.* **1998**, *108*, 564−578.

(5) Bettens, R. P. A.; Collins, M. A. Learning to interpolate molecular potential energy surfaces with confidence: A Bayesian approach. *J. Chem. Phys.* **1999**, *111*, 816−826.

(6) Collins, M. A.; Zhang, D. H. Application of interpolated potential energy surfaces to quantum reactive scattering. *J. Chem. Phys.* **1999**, *111*, 9924−9931.

(7) Boothroyd, A. I.; Keogh, W. J.; Martin, P. G.; Peterson, M. R. A refined H3 potential energy surface. *J. Chem. Phys.* **1996**, *104*, 7139−7152.

(8) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577−606.

(9) Bowman, J. M.; Czako, G.; Fu, B. High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094−8111.

(10) Chen, J.; Xu, X.; Xu, X.; Zhang, D. H. A global potential energy surface for the H2 + OH ↔ H2O + H reaction using neural networks. *J. Chem. Phys.* **2013**, *138*, 154301.

(11) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(12) Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. Combining ab initio computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules. *J. Chem. Phys.* **1996**, *105*, 7597−7604.

(13) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114*, 3371−3383.

(14) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930−17955.

(15) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits. *J. Phys. Chem. A* **2006**, *110*, 5295−5304.

(16) Manzhos, S.; Yamashita, K. A model for the dissociative adsorption of N2O on Cu(100) using a continuous potential energy surface. *Surf. Sci.* **2010**, *604*, 555−561.

(17) Manzhos, S.; Dawes, R.; Carrington, T. Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces. *Int. J. Quantum Chem.* **2015**, *115*, 1012−1020.

(18) Liu, T.; Zhang, Z.; Fu, B.; Yang, X.; Zhang, D. H. A seven-dimensional quantum dynamics study of the dissociative chemisorption of H2O on Cu(111): effects of azimuthal angles and azimuthal angle-averaging. *Chem. Sci.* **2016**, *7*, 1840−1845.

(19) Shen, X.; Chen, J.; Zhang, Z.; Shao, K.; Zhang, D. H. Methane dissociation on Ni(111): A fifteen-dimensional potential energy surface using neural network method. *J. Chem. Phys.* **2015**, *143*, 144701.

(20) Zhang, Z.; Liu, T.; Fu, B.; Yang, X.; Zhang, D. H. First-principles quantum dynamical theory for the dissociative chemisorption of H2O on rigid Cu (111). *Nat. Commun.* **2016**, *7*, 11953.

(21) Jiang, B.; Li, J.; Guo, H. Potential energy surfaces from high fidelity fitting of ab initio points: the permutation invariant polynomial - neural network approach. *Int. Rev. Phys. Chem.* **2016**, *35*, 479−506.

(22) Kondati Natarajan, S.; Morawietz, T.; Behler, J. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8356−8371.

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147

(23) Xu, X.; Chen, J.; Zhang, D. H. Global Potential Energy Surface for the H+CH4 ↔ H2+CH3 Reaction using Neural Networks. *Chin. J. Chem. Phys.* **2014**, *27*, 373−379.
(24) Cui, J.; Krems, R. V. Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes. *J. Phys. B: At., Mol. Opt. Phys.* **2016**, *49*, 224001.
(25) Kolb, B.; Marshall, P.; Zhao, B.; Jiang, B.; Guo, H. Representing Global Reactive Potential Energy Surfaces Using Gaussian Processes. *J. Phys. Chem. A* **2017**, *121*, 2552−2557.
(26) Guan, Y.; Yang, S.; Zhang, D. H. Construction of reactive potential energy surfaces with Gaussian process regression: active data selection. *Mol. Phys.* **2017**, *0*, 1−12.
(27) Rasmussen, C. E.; Williams, C. K. *Gaussian processes for machine learning*; MIT Press: Cambridge, MA, 2006; Vol. *1*.
(28) Paciorek, C. J.; Schervish, M. J. Nonstationary covariance functions for Gaussian process regression. *Advances in neural information processing systems*; **2004**; pp 273−280.
(29) Shao, K.; Fu, B.; Zhang, D. H. A global full-dimensional potential energy surface and quasiclassical trajectory study of the O(1D) + CH4 multichannel reaction. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24098−24107.
(30) Gramacy, R. B.; Lee, H. K. H. Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *J. Am. Stat. Assoc.* **2008**, *103*, 1119−1130.
(31) Sammut, C., Webb, G. I., Eds.; *Encyclopedia of Machine Learning*; Springer US: Boston, MA, 2010; pp 180−180.
(32) Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting potential energy surfaces with fundamental invariant neural network. *J. Chem. Phys.* **2016**, *145*, 071101.
(33) Hagan, M. T.; Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* **1994**, *5*, 989−993.
(34) Sarle, W. S. Stopped training and other remedies for overfitting. *Computing science and statistics* **1996**, 352−360.
(35) Agrafiotis, D. K.; Cedeño, W.; Lobanov, V. S. On the Use of Neural Network Ensembles in QSAR and QSPR. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 903−911.
(36) Zhou, Z.-H.; Wu, J.; Tang, W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence* **2002**, *137*, 239−263.
(37) Li, J.; Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems. *J. Chem. Phys.* **2013**, *139*, 204103.
(38) Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. 2007; pp 1027−1035.
(39) Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data*; John Wiley & Sons, Inc.: New York, 2008; pp 126−163.
(40) Ng, R. T.; Han, J. CLARANS: a method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering* **2002**, *14*, 1003−1016.
(41) Dueck, D. *Affinity propagation: clustering data by passing messages*; University of Toronto: Toronto, 2009.
(42) Frey, B. J.; Dueck, D. Clustering by passing Messages Between Data Points. *Science* **2007**, *315*, 972−976.

DOI: 10.1021/acs.jpca.8b00859
J. Phys. Chem. A 2018, 122, 3140−3147