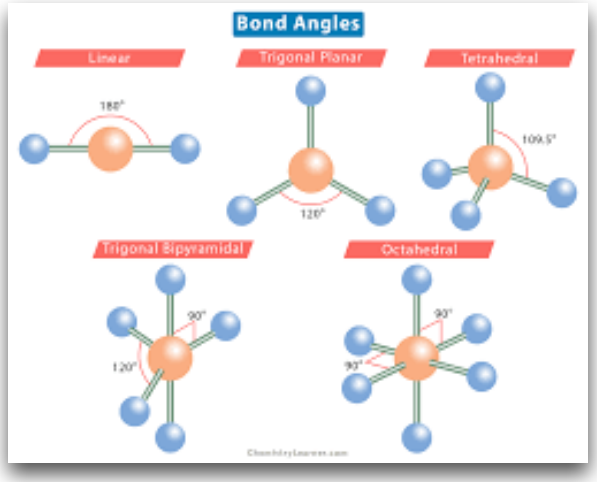
Global features

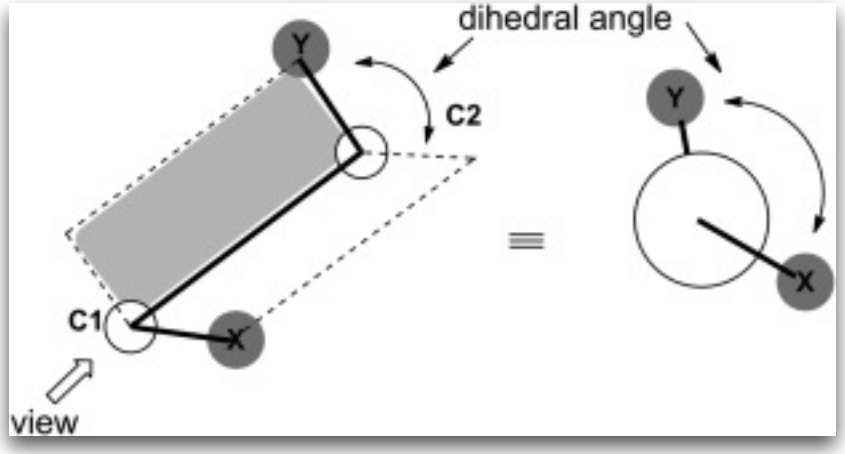
REPEAT

1. Many-body tensor representation

- Ewald matrix, sine matrix contain information about interatomic distances, but not bond angles or dihedrals.
- MBTR captures such information.
- Identify and encode information about motifs of different size, then group them according to their chemical composition.



https://www.chemistrylearner.com/bond-angles.html



https://www.sciencedirect.com/topics/chemistry/dihedral-angle

Arxiv:1904.08875

- Steps to construct MBTR

1. Identify structural motifs of k atoms involving specific chemical elements

(I) $k = 1, 2, 3$ etc.

(II) $k = 1$ motif for H or He or Fe etc. $k = 2$ motifs for H-M, M-O etc. $k = 3$ motifs: M-O-M, O-M-O, M-M-M etc.

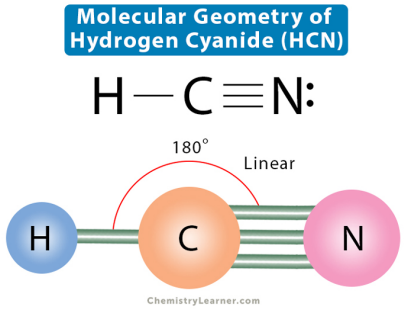
2. Map motif to a scalar function $g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k)$.

(I) Commonly used functions:

(II) $g_1^{Z_i}(\vec{R}_i) = Z_i$, one for H, one of C, … one for each element present

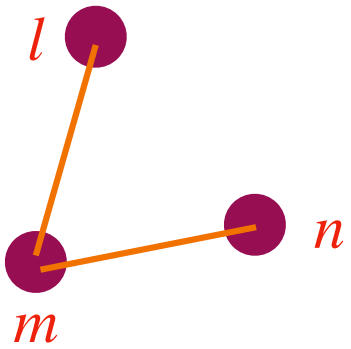
(III) $g_2(\vec{R}_i, \vec{R}_j) = \dfrac{1}{|\vec{R}_i - \vec{R}_j|}$, Z indices suppressed for now.

(IV) $g_3(\vec{R}_l, \vec{R}_m, \vec{R}_n) = \cos(\angle(\vec{R}_l - \vec{R}_m, \vec{R}_m - \vec{R}_n))$.



3. $g_k's$ are 'broadened' using a (Gaussian) kernel

(I) $g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k) \rightarrow \mathscr{D}_k(x, g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k))$



Motif $\rightarrow g_k's \rightarrow \mathscr{D}_k$'s

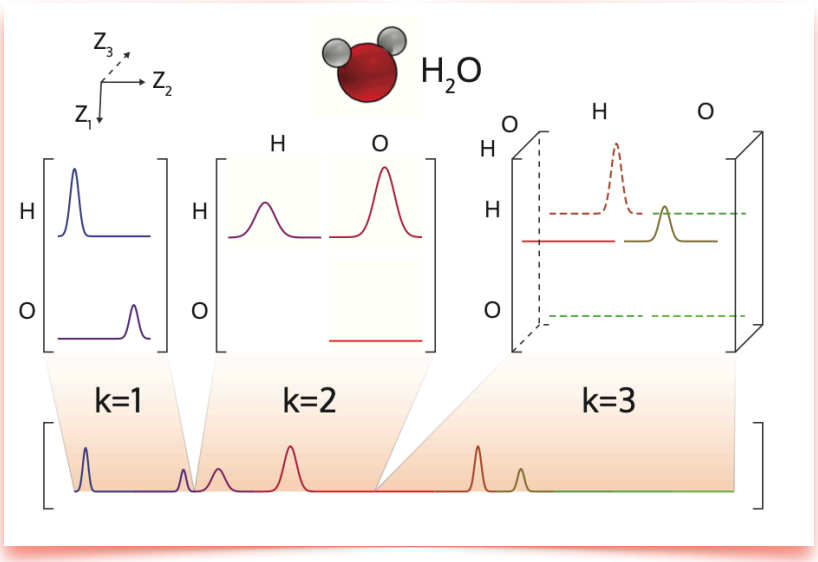
- More details of MBTR:
  - $g_1(\vec{R}_i) = Z_i, \quad g_2(\vec{R}_i, \vec{R}_j) = \frac{1}{|\vec{R}_i - \vec{R}_j|}.$
  - $g_3(\vec{R}_l, \vec{R}_m, \vec{R}_n) = \cos(\angle(\vec{R}_l - \vec{R}_m, \vec{R}_m - \vec{R}_n)).$
  - $\mathscr{D}_1^i(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp(-\frac{(x - g_1(\vec{R}_i))^2}{2\sigma_1^2}),$
  - $\mathscr{D}_2^{i,j}(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp(-\frac{(x - g_2(\vec{R}_i, \vec{R}_j))^2}{2\sigma_2^2}),$ etc.

Arxiv:1904.08875

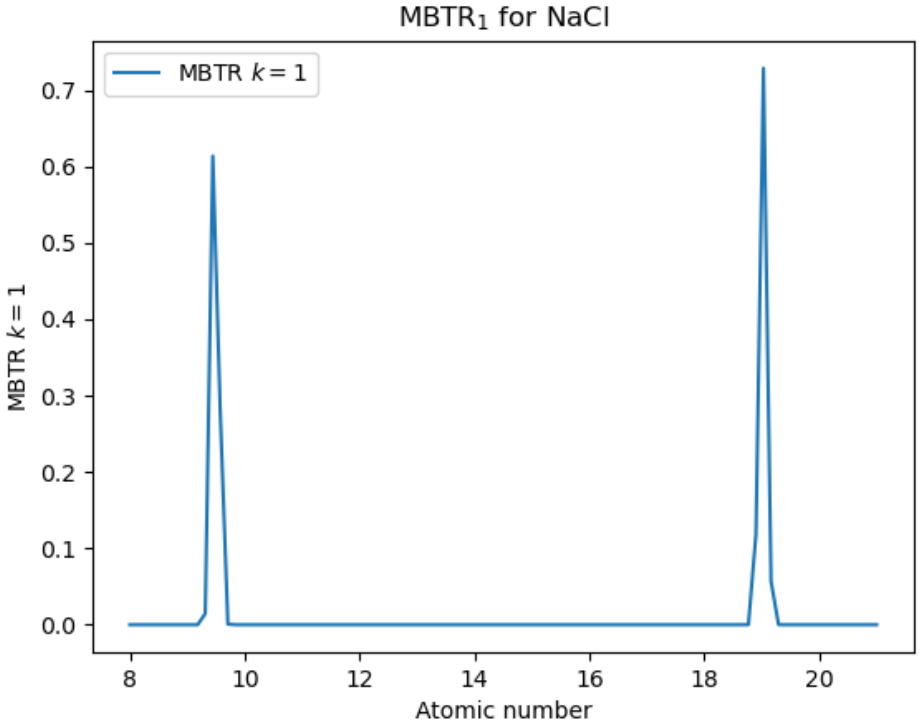
- Now f_k's are the MBTRs
  - $f_1^{Z_1}(x) = \sum_i^{N_e} w_1^i \mathscr{D}_1(x) \delta_{Z_i,Z_1},$
  - $f_2^{Z_1,Z_2}(x) = \sum_{i,j}^{N_e} w_2^{i,j} \mathscr{D}_2^{i,j}(x) \delta_{Z_i,Z_1} \delta_{Z_j,Z_2},$
  - $f_3^{Z_1,Z_2,Z_3}(x) = \sum_{l,m,n}^{N_e} w_2^{l,m,n} \mathscr{D}_2^{i,j}(x) \delta_{Z_l,Z_1} \delta_{Z_m,Z_2} \delta_{Z_n,Z_3}.$

- MBTR continued:
  - Weights
  - $w_1^i = 1 \quad \forall i,$
  - $w_2^{i,j} = e^{-s_2|\vec{R}_i - \vec{R}_j|},$
  - $w_3^{l,m,n} = e^{-s_3(|\vec{R}_l - \vec{R}_m| + |\vec{R}_m - \vec{R}_n| + |\vec{R}_n - \vec{R}_l|)}.$
  - Exponential forms ensure that contributions of motifs of distant atoms are damped out.

MBTR of H2O molecule

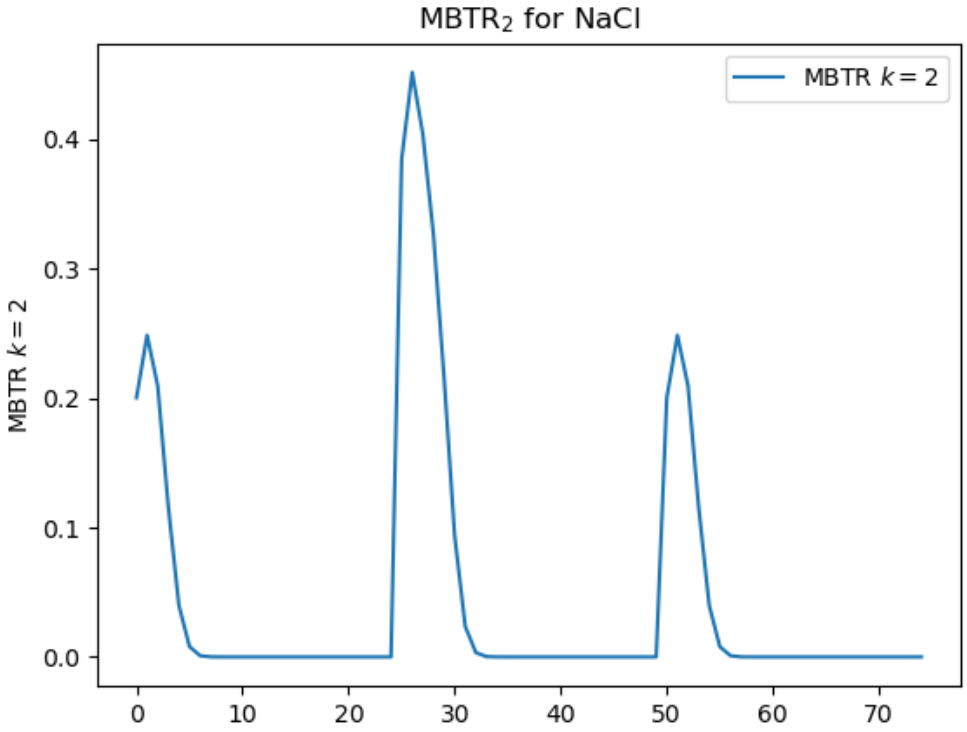


Arxiv:1904.08875



MBTR1 for NaCl
Vary the parameters to see if you can get the peaks at 11 and 17.

MBTR2 for NaCl with g_2 as inverse distance

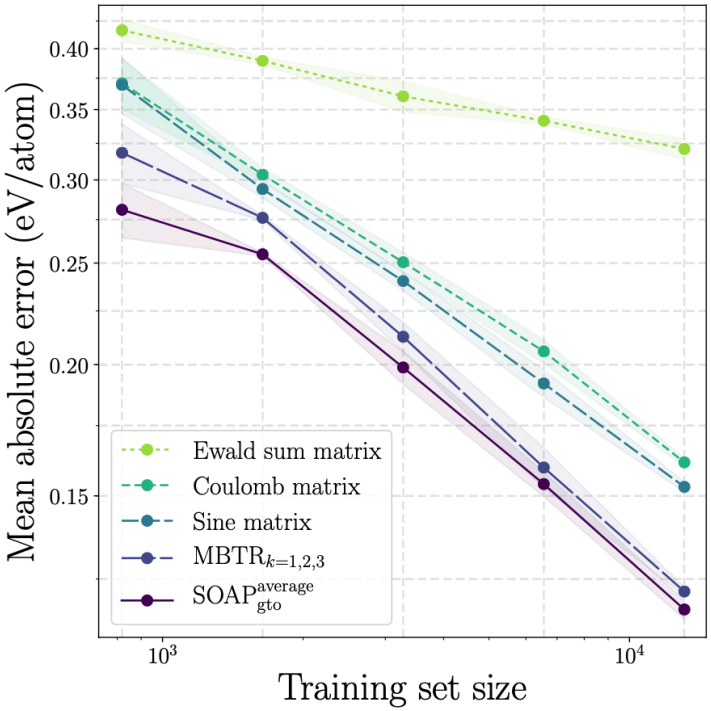


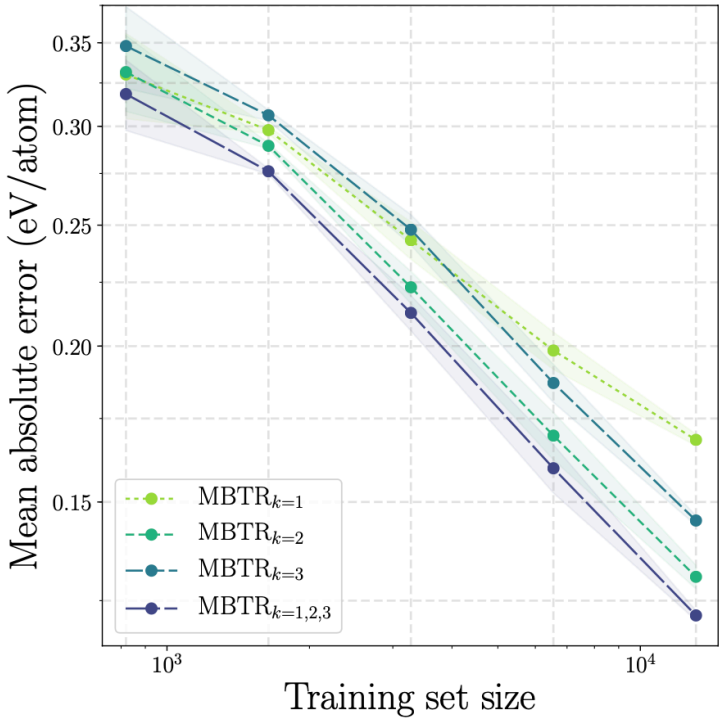
Task

1. Construct MBTR2 for NaCl with g_2=distance
2. Calculate the interatomic distances by hand and see if you can understand the peak locations

Performance of features for h_{form} predictions

- Data from OQMD database
  - Max 10 atoms per unit cell
  - Max 6 different elements
  - 222,215 samples after removing unconverged structures
  - KRR model trained to predict h_{form} for different feature vectors
  - Results:





Arxiv:1904.08875