ML-based prediction of partial atomic charges

Example 1: Pfizer Charge Assignment Method (PCAM) \[1\]

- An atom’s PCAM charge depends only on its neighbors’ configuration;
- It is limited to H, C, N, O, F, S, Cl atoms;
- Each element has a charge model that uses **radical and angular symmetry functions** to represent local surroundings;
- It employs random forest ML-method and the charge is calculated by: $q_j^0 = \sum_{i} T_i(A_j)$, where B is the tree count;
- The formula calculates corrected charges: $q_j = q_j^0 - \frac{\sum q_j^0 \cdot \sigma_j \cdot \Delta Q}{Q_{abs}}$.

Example 3: Atom-centered Atom Pairs (AP) fingerprint \[3\]

- Partial charges were derived using TPSSh/def2-TZVP electron densities and DDEC method;
- AP fingerprints (atom type, number of heavy-atom neighbors, distances, etc.) represent atoms;
- A **2D topological representation** eliminates conformational dependency and ensures transferability;
- It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms;
- RF generates accurate predictions of unseen compounds.

Example 2: ContraDRG \[2\]

- Molecules are represented as **cyclic undirected graphs** with additional information like the number of nested circles, distances between nearby atoms, etc.;
- Atomic positions and adjacency matrix (bonds) are needed to encode molecules;
- It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms;
- Each element has a charge model;
- A detailed benchmark shows KR is the best model.

Example 4: Atom-Path-descriptor (APD) \[4\]

- Using APD descriptors (**2 and 3D information**) to predict partial atomic charges;
- Atoms are represented by multiple layers, with each layer encoding information such as atom type, distance, bond type, and so on.

<table>
<thead>
<tr>
<th>Element</th>
<th>RMSE [e*10^-3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>93  72  63</td>
</tr>
<tr>
<td>P</td>
<td>685 561 372</td>
</tr>
<tr>
<td>C</td>
<td>220 142 114</td>
</tr>
</tbody>
</table>

Example 5: Multiple-Choice Knapsack Problem (\(\varepsilon\)-MCKP) \[5\]

- Molecules are considered as **graphs** (G = (V, E, t));
- Create a database with a record for each isomorphism class in the subgraph list for each atom;
- Each isomorphism class's partial charges are condensed into a histogram.