A Gaussian Process Library for Molecules

- FlowMO is a Gaussian Process library for molecules
- Representations include SMILES and ECFP6 fingerprints
- Bespoke kernels for these representations include string kernels for SMILES and the Tanimoto kernel for fingerprints.

Regression Benchmark

- We benchmark regression performance on three small molecular datasets: The Photoswitch dataset [1], ESOL [2] and FreeSolv [3].
- Compare against the best reported model from the MoleculeNet benchmark in addition to recently reported SOTA models.
- Achieve best performance on the photoswitch dataset

Table 1: RMSE of the models across the three datasets, with the scores of the best GP model and best overall model highlighted.

<table>
<thead>
<tr>
<th>Model</th>
<th>Photoswitch</th>
<th>ESOL</th>
<th>FreeSolv</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSK GP (SMILES)</td>
<td>26.0 ± 3.6</td>
<td>0.65 ± 0.01</td>
<td>1.29 ± 0.22</td>
</tr>
<tr>
<td>TK GP (Fingerprints)</td>
<td>22.6 ± 4.0</td>
<td>0.98 ± 0.08</td>
<td>1.85 ± 0.10</td>
</tr>
<tr>
<td>ANP</td>
<td>27.2 ± 3.7</td>
<td>1.32 ± 0.13</td>
<td>2.05 ± 0.47</td>
</tr>
<tr>
<td>BNN</td>
<td>25.5 ± 5.0</td>
<td>1.01 ± 0.11</td>
<td>1.92 ± 0.20</td>
</tr>
<tr>
<td>MoleculeNet</td>
<td>22.0 ± 3.5</td>
<td>0.58 ± 0.03</td>
<td>1.15 ± 0.02</td>
</tr>
<tr>
<td>SMILES-X</td>
<td>0.70 ± 0.05</td>
<td>1.14 ± 0.17</td>
<td></td>
</tr>
<tr>
<td>SMILES-X (Augm)</td>
<td>0.57 ± 0.07</td>
<td>0.81 ± 0.22</td>
<td></td>
</tr>
</tbody>
</table>

Uncertainty Calibration

To analyse the calibration achieved by the predictive distributions provided by the probabilistic models (only the GPs, BNN and ANP), we define a calibration score function

\[
C(q) = \frac{1}{|T|} \sum_{m \in T} \mathbb{1}\left( \left| \frac{\hat{y}(m) - y(m)}{\hat{\sigma}(m)} \right| < \Phi^{-1}\left(1 + \frac{q}{2}\right) \right)
\]

based on cross-validatory predictive p-values. \( y(m), \hat{y}(m) \) and \( \hat{\sigma}(m) \) represent true values, predictive means and predictive standard deviations for each test molecule \( m \in T \), and \( \Phi^{-1} \) is the inverse of the standard Gaussian cumulative distribution function. The indicator \( 1 \) is activated only when the true value is contained in the model’s \( q \times 100\% \) confidence interval. Therefore, perfect calibration at the \( q \)th quantile corresponds to \( C(q) = q \). \( C(q) > q \) indicates under-confidence through overly large uncertainty estimates (limiting the strength of conclusions that can be drawn from the model) whereas \( C(q) < q \) denotes over-confidence (leading to reckless decisions downstream). We plot \( C(q) \) for our probabilistic models in Figure 2.

Uncertainty Calibration Benchmark

String kernel GP demonstrates superior calibration across all tasks.

Future Work

We plan to extend the library to cater for graph representations of molecules by making use of graph kernel GPs.

References