NEURAL INFORMATION PROCESSING SYSTEMS

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF OXFORD

Stanford University

EPFL

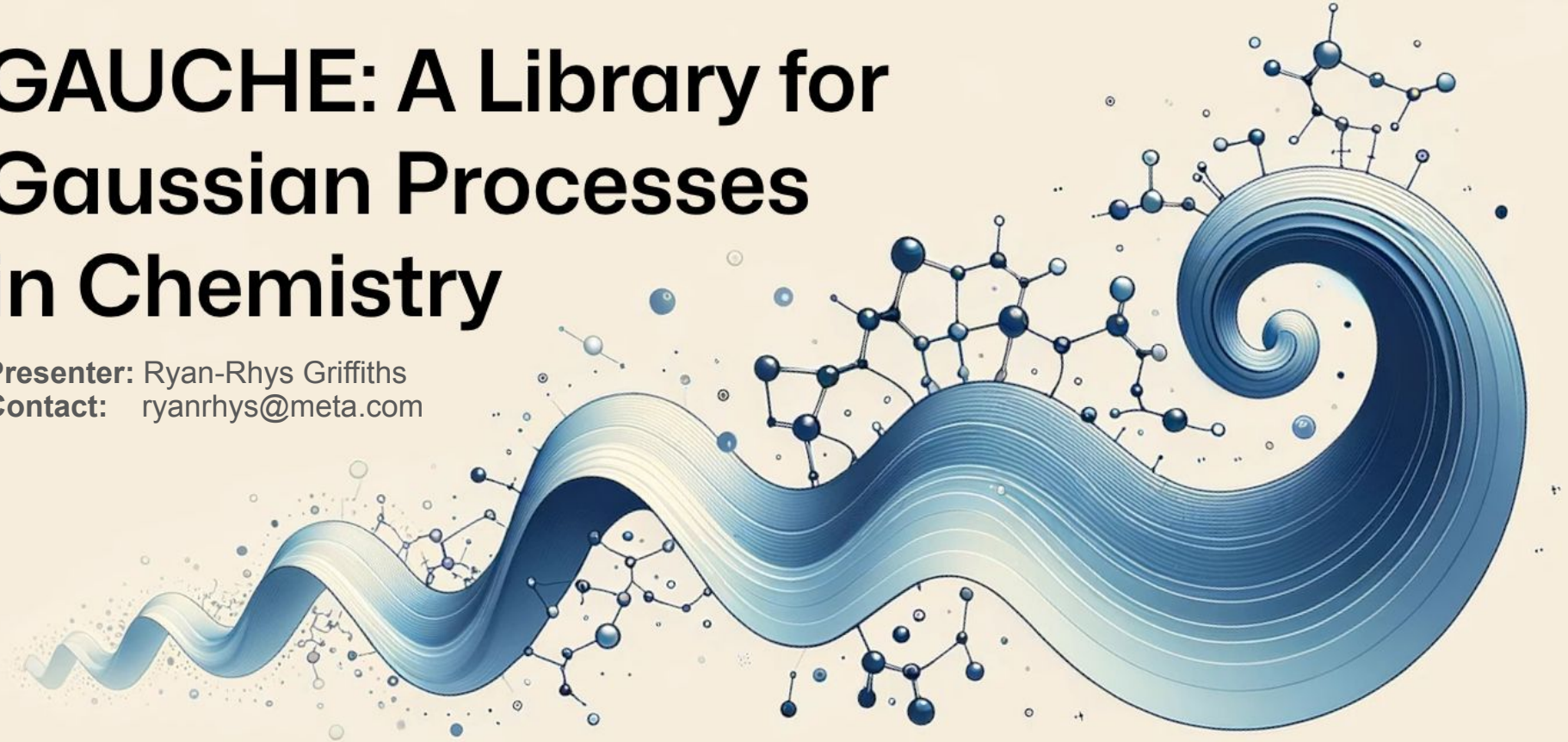
UNIVERSITY OF TORONTO

Cornell University

Mila

GAUCHE: A Library for Gaussian Processes in Chemistry

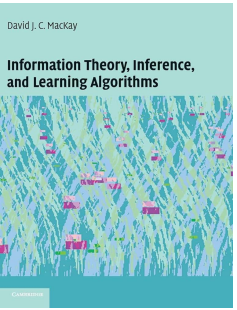
Presenter: Ryan-Rhys Griffiths
Contact: ryanrhys@meta.com



Gaussian processes (GPs) vs. Deep Neural Networks (DNNs)

Model	Capability	Check
DNN	Variants can operate on molecular representations	
GP	Variants have difficulty operating on many molecular representations	
GP	Exact Bayesian inference. First choice surrogate for Bayesian optimisation	
DNN	Approximate Bayesian inference. Challenging to use for Bayesian optimisation	

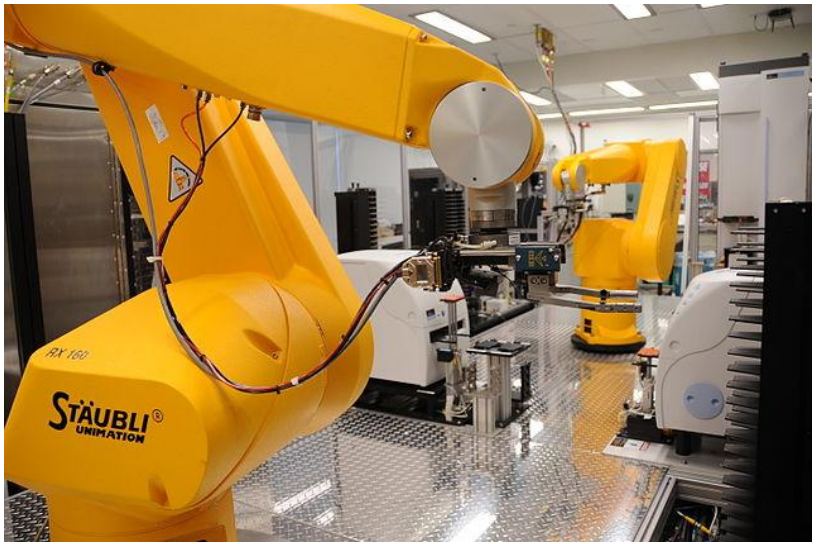
Gaussian Processes as Stable Surrogate Models for Active Learning and Bayesian Optimisation Loops



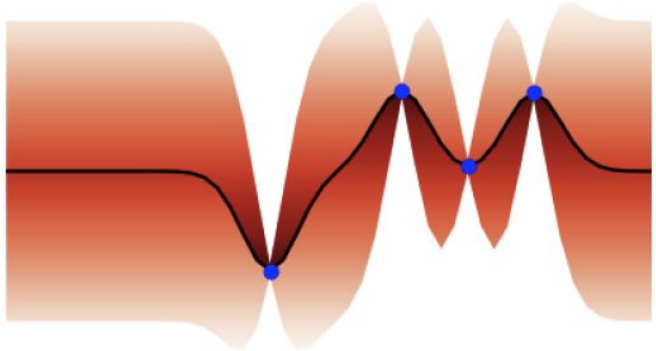
"Gaussian processes are useful tools for automated tasks where fine tuning for each problem is not possible. We do not appear to sacrifice any performance for this simplicity."

Sir David MacKay FRS

Gaussian Processes as a Tool for Self-Driving Laboratories

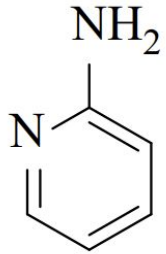


Extending Gaussian Processes to Molecular Representations



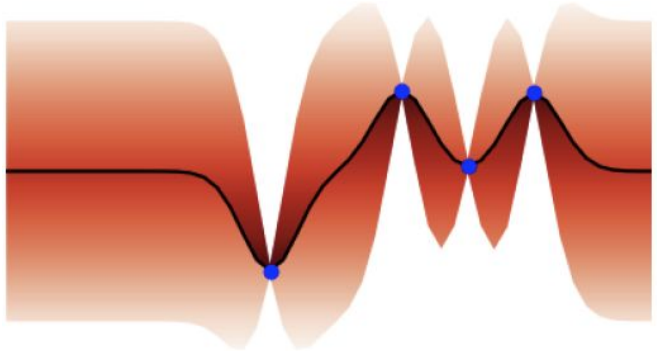
?

NC1=CC=CC=N1



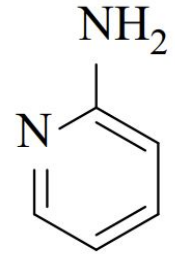
$$\begin{bmatrix} 1 & 0 & \cdots & 1 \end{bmatrix}^{\top}$$

Extending Gaussian Processes to Molecular Representations



String, Graph, and Bit Vector Kernels

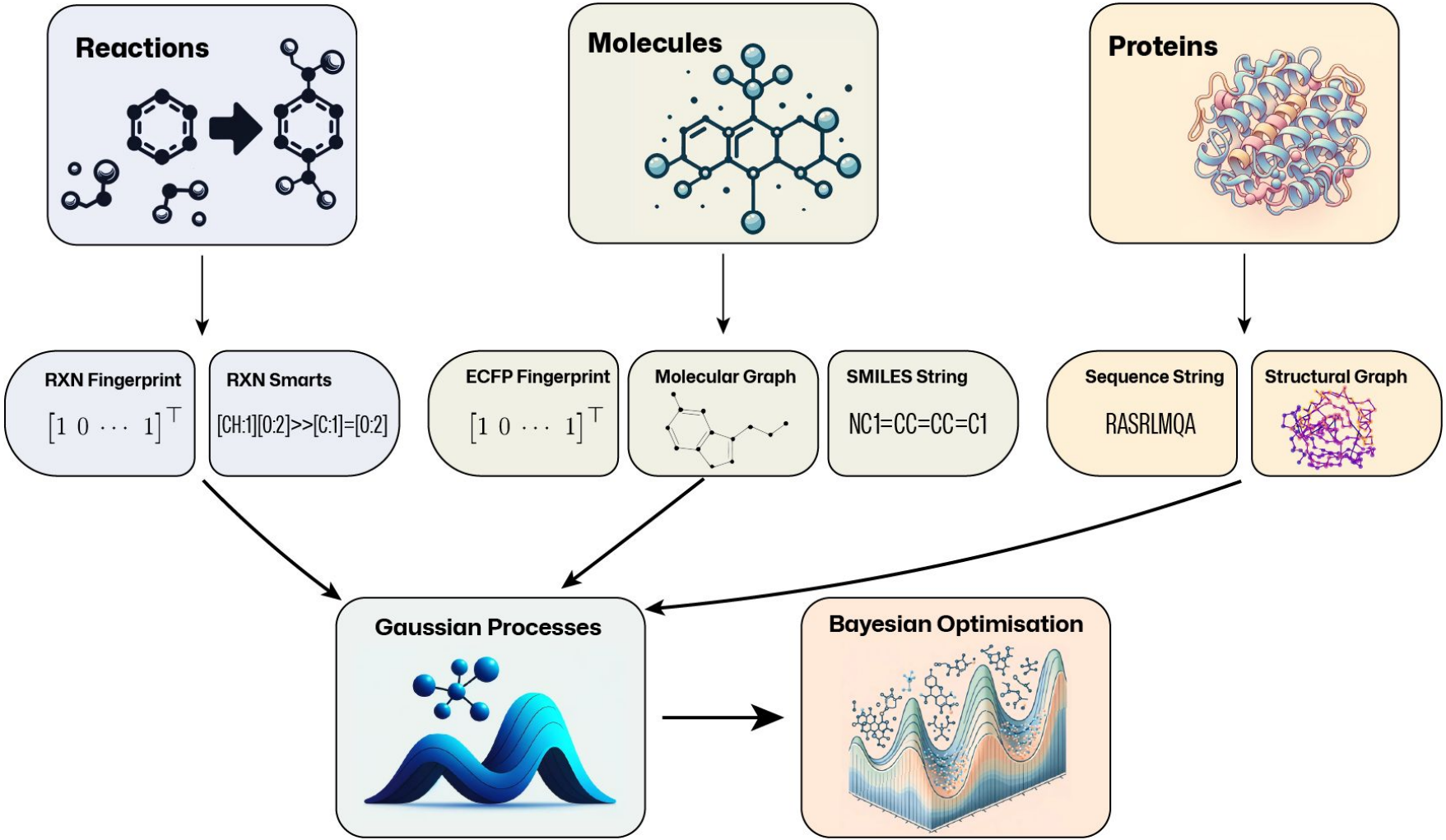
NC1=CC=CC=N1



$$\begin{bmatrix} 1 & 0 & \cdots & 1 \end{bmatrix}^\top$$

Comparison to Other Software Libraries

Library	Gaussian Processes	Bayesian Optimisation	Molecular Representations	Chemistry Tutorials	Graph Kernels	Bit Vector Kernels	String Kernels
GPyTorch [26]	✓	✗	✗	✗	✗	✗	✗
GPflow [69, 70]	✓	✗	✗	✗	✗	✗	✗
BoTorch [27]	✓	✓	✗	✗	✗	✗	✗
DeepChem [71]	✗	✗	✓	✓	✗	✗	✗
GraKel [59]	✗	✗	✗	✗	✓	✗	✗
FlowMO [72]	✓	✗	✓	✓	✗	✓	✓
GAUCHE (ours)	✓	✓	✓	✓	✓	✓	✓


Reactions
Molecules
Proteins
RXN Fingerprint
$[1\ 0\ \cdots\ 1]^\top$
RXN Smarts
[CH:1][0:2]>>[C:1]=[0:2]
ECFP Fingerprint
$[1\ 0\ \cdots\ 1]^\top$
Molecular Graph
SMILES String
NC1=CC=CC=C1
Sequence String
RASRLMQA
Structural Graph
Gaussian Processes
Bayesian Optimisation

Tutorials

```python
import gpytorch
from botorch import fit_gpytorch_model
from gauche.kernels.fingerprint_kernels.tanimoto_kernel import TanimotoKernel

class TanimotoGP(gpytorch.models.ExactGP):
  def __init__(self, train_x, train_y, likelihood):
    super(TanimotoGP, self).__init__(train_x, train_y, likelihood)
    self.mean_module = gpytorch.means.ConstantMean()
    self.covar_module = gpytorch.kernels.ScaleKernel(TanimotoKernel())

  def forward(self, x):
    mean_x = self.mean_module(x)
    covar_x = self.covar_module(x)
    return gpytorch.distributions.MultivariateNormal(mean_x, covar_x)

# initialise GP likelihood, model and
# marginal log likelihood objective
likelihood = gpytorch.likelihoods.GaussianLikelihood()
model = TanimotoGP(X_train, y_train, likelihood)
mll = gpytorch.mlls.ExactMarginalLogLikelihood(likelihood, model)

# fit GP with BoTorch in order to use
# the LBFGS-B optimiser (recommended)
fit_gpytorch_model(mll)

# use the trained GP to get predictions and
# uncertainty estimates for new molecules
model.eval()
likelihood.eval()
preds = model(X_test)
pred_means, pred_vars = preds.mean, preds.variance
```

gauche

Public

Unwatch 5 Fork 11 Starred 108

main 21 branches 0 tags

Go to file Add file Code

leojklarner Polished README and included new overview plot. cf61f4d 2 weeks ago 259 commits

.github	Add issue / PR templates	last year
.requirements	Refactor setup to support modular requirements lists.	last year
benchmarks	Added multioutput GP tutorial. Refactor repo name to gauche.	last year
data	Polished README and included new overview plot.	2 weeks ago
docs	Update index.rst	2 months ago
gauche	Adding:	3 months ago
imgs	Polished README and included new overview plot.	2 weeks ago
notebooks	Merge pull request #58 from Ryan-Rhys/ryan/dev	3 months ago
tests	Adding:	3 months ago
.gitignore	add ssk experiment	8 months ago
.pre-commit-config.yaml	Add pre-commit config	last year
LICENSE	Anonymised name on license	last year
Makefile	add black & isort	3 years ago
README.md	Polished README and included new overview plot.	2 weeks ago
citation.bib	Create citation.bib	10 months ago
conda_env.yml	add rxnfp dependency for the protein_ssk.ipynb	8 months ago
pyproject.toml	add black & isort	3 years ago
setup.py	Ensure all requirements are list	last year

README.md

repo status Active License MIT docs passing codefactor A code style black launch binder

GAUCHE is a collaborative, open-source software library that aims to make state-of-the-art probabilistic modelling and black-box optimisation techniques more easily accessible to scientific experts in chemistry, materials science and beyond. We provide 30+ bespoke kernels for molecules, chemical reactions and proteins and illustrate how they can be used for Gaussian processes and Bayesian optimisation in 10+ easy-to-adapt tutorial notebooks.

Overview | Getting Started | Documentation | Paper (NeurIPS 2023)

About

A Library for Gaussian Processes in Chemistry

leojklarner.github.io/gauche/

Readme
MIT license
Cite this repository
Activity
108 stars
5 watching
11 forks
Report repository

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 10

Deployments 15

github-pages 2 months ago

+ 14 deployments

Languages

Jupyter Notebook 75.4%
Python 24.5%
Other 0.1%