Popular science summary of the PhD thesis

PhD student	Maximillian Fornitz Vording
Title of the PhD thesis	Gaussian Processes for Modeling Physical Systems
PhD school/Department	Mathematics, Physics and Informatics

Science summary

* Please give a short popular summary in Danish or English (approximately half a page) suited for the publication of the title, main content, results and innovations of the PhD thesis also including prospective utilizations hereof. The summary should be written for the general public interested in science and technology:

The space of possible molecules is vast and almost unexplored. With an upper bound on the number of possible molecules being 10^{180} , which is twice the magnitude of the estimated number of atoms in the universe, 10^{80}, we quickly run out of time, resources and PhD students to study the properties of every single molecule in the lab or via time-consuming simulations in quantum chemistry. To guide us through chemical space to the potential candidates for new materials and drugs, we need data-driven machine learning methods for predicting the properties of molecules in unexplored regions with uncertainties.

Deep neural network models have accelerated the field of molecular property prediction via graph convolutions since 2015, but they come with two caveats: Firstly, they do not easily provide us with calibrated uncertainties on the predictions. Secondly, they are prone to overfitting small amounts of training data. Both caveats make them less suitable as tools in Bayesian optimisation for automatic experimental design in screening of molecular properties. Therefore this thesis investigates the use of Gaussian processes as a better alternative for this, since they provide uncertainty estimates and smooth interpolation via the structure in the kernel.

Two methods and applications are considered for this: The first, is a warped Gaussian process regression to track and infer peak shapes through the low signal to noise regions of resonance spectra from a thermomechanical analysis experiment used to characterise the properties of a crystallised drug. The second, is a novel graph convolutional Gaussian process and a message passing neural network extended to work on rotation invariant internal degrees of freedom in molecules, to predict the solubility of them in two experimental datasets. We compare them on learning, calibration and per-molecule-point error curves, for understanding their ability to learn from the least amount of available training data. This method shows great promise for small amounts of training data.

Popular science summary of the PhD thesis

PhD student	Maximillian Fornitz Vording

Please email the summary to the PhD secretary at the department