Generative models for structure-based drug design

David Koes

@david_koes

Revolutionizing Chemistry with Artificial Intelligence
American Chemical Society
Boston, MA
August 20, 2018
Machine Learning

Features $X$ → Model → Prediction $y$
Neural Networks

The universal approximation theorem states that, under reasonable assumptions, a feedforward neural network with a finite number of nodes can approximate any continuous function to within a given error over a bounded input domain.
Deep Learning

\[ \delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l) \]

\[ \frac{\partial L}{\partial w^l_{jk}} = a_k^{l-1} \delta^l_j \quad \text{and} \quad \frac{\partial L}{\partial b^l_j} = \delta^l_j \]
Discriminative Model

Features $X$ → Prediction $y$
Generative Model

Features $X \rightarrow y$?
Generative Adversarial Networks

- True Examples
- Generator
- Discriminator
- Loss
  Is this a real dog picture?
Generative Adversarial Networks

https://arxiv.org › stat ▼
by IJ Goodfellow - 2014 - Cited by 4339 - Related articles
Jun 10, 2014 - Submission history. From: Ian Goodfellow [view email] [v1] Tue, 10 Jun 2014 18:58:17 GMT (1257kb,D). Which authors of this paper are ...

http://torch.ch/blog/2015/11/13/gan.html
PROGRESSIVE GROWING OF GANs FOR IMPROVED QUALITY, STABILITY, AND VARIATION

Tero Karras
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

Jaakko Lehtinen
NVIDIA
Aalto University

https://youtu.be/G06dEcZ-QTg
Generative models for structure-based drug design
Structure Based Drug Design

Pose Prediction  Binding Discrimination  Affinity Prediction

Virtual Screening  Lead Optimization
Drug Discovery Funnel

http://pharmit.csb.pitt.edu

COMP 528: Structure-based searching of chemical space with Pharmit
Thursday, Aug 23 9:20 AM
Douglass, Westin Boston Waterfront
Drug Discovery Funnel

Matching

Scoring

Dynamics

\[ \text{gauss}_i(d) = \frac{w_{\text{gauss}_i} \cdot e^{-(d/0.5)^2}}{d} \]

\[ \text{gauss}_o(d) = \frac{w_{\text{gauss}_o} \cdot e^{-(d/0.5)^2}}{d} \]

\[ \text{repulsion}(d) = \begin{cases} w_{\text{repulsion}}d^2 & d < 0 \\ 0 & d \geq 0 \end{cases} \]

\[ \text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & \text{otherwise} \end{cases} \]

\[ \text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(-0.7/d) & \text{otherwise} \end{cases} \]

Protein-Ligand Scoring

Model

Pose Prediction
Binding Discrimination
Affinity Prediction
Convolutional Neural Networks

Convolution

Feature Maps

Fully Connected

Traditional NN

Dog: 0.99
Cat: 0.02
Convolutional Filters

The filters shown are used for image processing. Each filter consists of a grid of numbers that are applied to an image to perform convolution operations. The numbers in each row represent how each pixel in the image is weighted when the filter is applied. For example, the filter on the left has the pattern 

\[
\begin{array}{ccc}
-1 & -1 & -1 \\
0 & 0 & 0 \\
1 & 1 & 1 \\
\end{array}
\]

These filters are typically used to detect edges or other features in images. The effect of these filters on an image is shown in the right column of the image.
Protein-Ligand Representation

(R,G,B) pixel → (Carbon, Nitrogen, Oxygen,...) voxel

The only parameters for this representation are the choice of grid resolution, atom density, and atom types.
Why Grids?

**Cons**
- coordinate frame dependent
- pairwise interactions not explicit

**Pros**
- clear spatial relationships
- amazingly parallel
- easy to interpret
Model

2x2x2 Max Pooling  2x2x2 Max Pooling  2x2x2 Max Pooling  2x2x2 Max Pooling  3x3x3 Convolution  3x3x3 Convolution  3x3x3 Convolution  3x3x3 Convolution

48x48x48x35  24x24x24x35  24x24x24x32  12x12x12x32  12x12x12x64  24x24x24x35  12x12x12x64  6x6x6x64

Affinity
Pose Score

Fully Connected  Fully Connected  Rectified Linear Unit  Rectified Linear Unit  Rectified Linear Unit  Rectified Linear Unit

Softmax+Logistic Loss  Pseudo-Huber Loss

Rectified Linear Unit  Rectified Linear Unit  Rectified Linear Unit  Rectified Linear Unit
Results

Spearman = 0.570, RMSE = 1.704

CNN (AUC=0.90)
Vina (AUC=0.63)
Beyond Scoring

\[
\frac{\partial L}{\partial w_{jk}^{l}} = a_{k}^{l-1} \delta_{j}^{l} \quad \text{and} \quad \frac{\partial L}{\partial b_{j}^{l}} = \delta_{j}^{l}
\]

Deep Dreams

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
Beyond Scoring

2Q89

Less Oxygen Here

More Oxygen Here

\[ \frac{\partial L}{\partial A} = \sum_{i \in G^A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A} \]
Iterative Refinement

- Best
- First Minimization
- Second Iteration
- Third Iteration

# Poses vs. RMSD Change
Generative Models

Generative models approximate a data distribution directly. They can map samples from one distribution (noise or input data) to realistic samples from an output distribution of interest.

noise sample

Generator

generated receptor & ligand grid
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Rafael Gómez-Bombarelli†, Jennifer N. Wei‡, David Duvenaud†, José Miguel Hernández-Lobato§, Benjamín Sánchez-Lengeling‡, Dennis Sheberla†, Jorge Aguilera-Iparraguirre†, Timothy D. Hirzel‡, Ryan P. Adams‡, and Alán Aspuru-Guzik***
Context Encoding

http://people.eecs.berkeley.edu/~pathak/context_encoder/
Context Encoding

receptor grid

Generator

generated ligand grid
Receptor-Conditional Ligand-Variational Model

L2 loss

GAN loss
Receptor-Conditional Ligand-Variational Model
Receptor-Conditional Ligand-Variational Model

GAN loss

Discriminator
Model Architecture

![Diagram of model architecture with parameters like n_levels, n_filters, convolution, pooling, width_factor, n_latent, and conv_per_level.](image-url)
Model Architecture

n_levels = 3
conv_per_level = 3
width_factor = 2
n_latent = 1024
Training Procedure

2016 PDBbind refined set
3765 crystal structures
Vina docking
RMSD < 2 Å from crystal pose
8648 poses (~2.3 per target)
random rotation & translation

Adam optimization
base_lr = 0.00001
momentum = 0.9
momentum2 = 0.999
max_iter = 100000
batch_size = 50
Autoencoding Examples
Autoencoding Examples
Autoencoding Examples
Atom Fitting

\[ a^* = \arg \min_a \|d - D(a)\|^2 + \lambda E(a) \]
Ligand Variation

to be continued…
Conditioning on the Receptor
Acknowledgements

Matt Ragoza

Jocelyn Sunseri  Paul Francoeur

AI GRANT  National Institute of General Medical Sciences R01GM108340  NVIDIA  Google Cloud
github.com/gnina
http://bits.csb.pitt.edu
@david_koes