

機械学習を用いた 機能分子の自動設計

津田宏治
東京大学/物材機構/理研AIP

Deep Learning: 予測から生成へ

Art Can Google's Deep Dream become an art machine?
The company's neural network has created a slew of beautiful and at times terrifying images, and is being harnessed to create unique artwork


1754 207

Alex Rayner

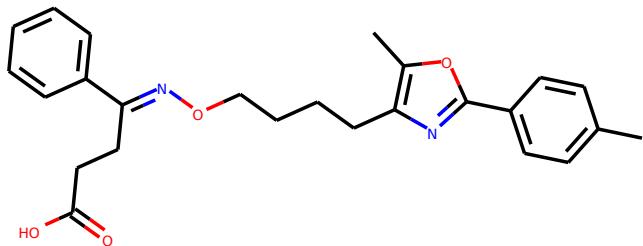
Monday 28 March 2016
22.51 BST



Moonage Daydream: art created by Deep Dream. Photograph: Deep Dream

新規化合物生成

機械学習(予測)



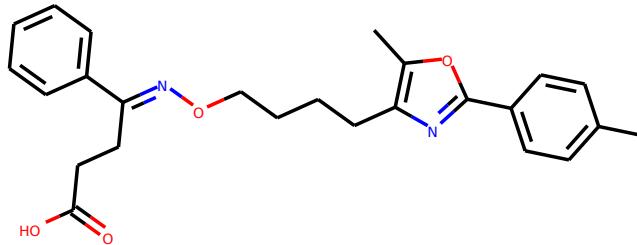
Properties:

Binding Score
Bio Activity
ADMET
...

望ましい性質&
正しい(存在可能な)化合物の生成

SMILES

- SMILESの例



Cc3ccc(c2nc(CCCCCO/N=C(CCC(O)=O)c1cccccc1)c(C)o2)cc3

- SMILESの要素

Atom: {C, c, o, O, N, F, [C@@H], n, -, S, Cl, [O-], [C@H], [NH+], [C@], s, Br, [nH], [NH3+], [NH2+], [C@@], [N+], [nH+], [S@], [N-], [n+], [S@@], [S-], I, [n-], P, [OH+], [NH-], [P@@H], [P@@], [PH2], [P@], [P+], [S+], [o+], [CH2-], [CH-], [SH+], [O+], [s+], [PH+], [PH], [S@@+] }

Bonds: {/, =, ¥# }

Ring: {1,2,3,4,5,6,7,8,9}

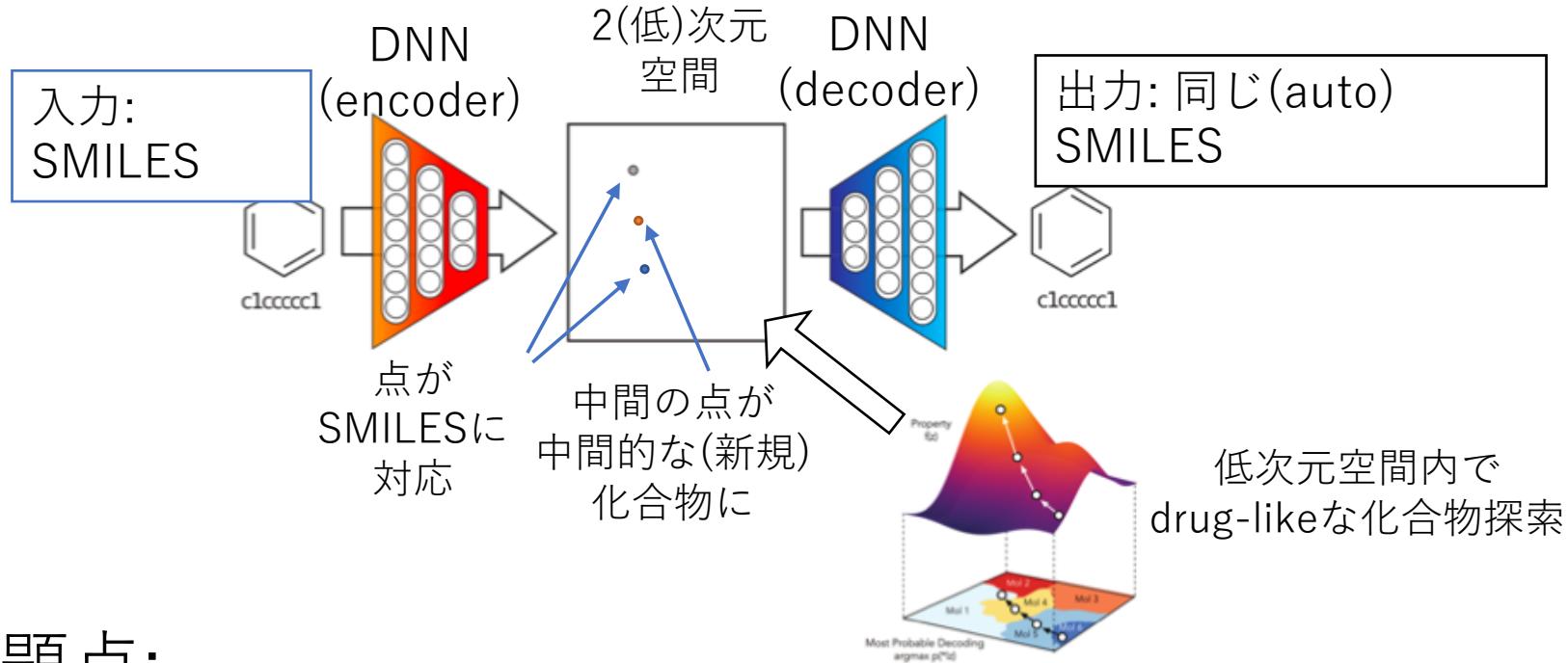
Branch: {(,)}

De novo 化合物生成

- これまでの化合物自動生成法は、あらかじめ定義されたFragmentを組み合わせるものが多かつた
- 深層学習によるde novo生成
 - Variational autoencoder (Kusner et al., ICML 2017)
 - Recurrent neural network (Segler et al., Arxiv, 2017)
- ChemTS (<https://github.com/tsudalab/ChemTS>)
 - モンテカルロ木探索 + Recurrent Neural Network

VAEに基づく化合物生成

- Variational Auto Encoder (VAE).



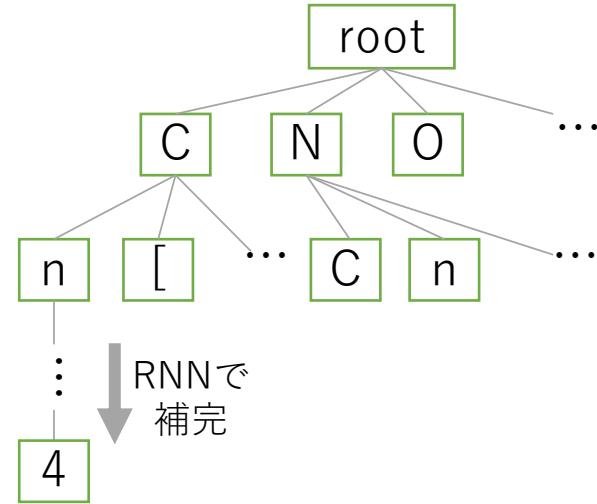
- 問題点:

- 正しいSMILESの生成確率が低い ($0.4\% \rightarrow 5\%$ 程度)
- 遅い

Gómez-Bombarelli et al., "Automatic chemical design using a data-driven continuous representation of molecules," Arxiv, 2016.
Kusner et al., "Grammar Variational Autoencoder," ICML, 2017.

ChemTS: MCTS and RNN

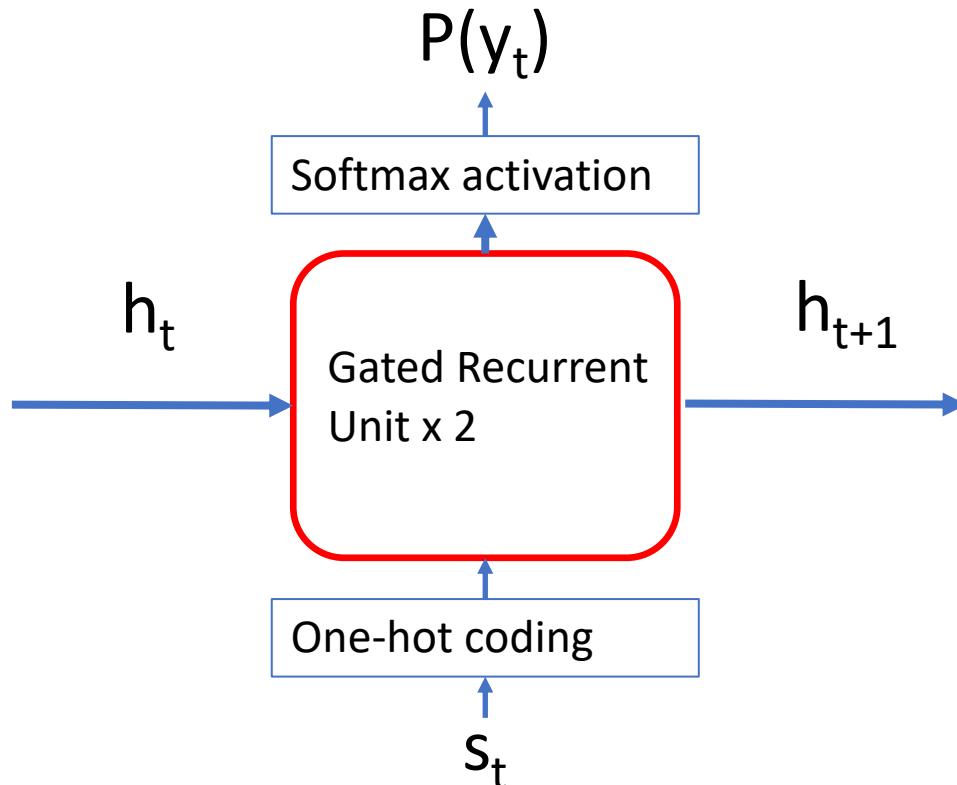
- SMILES文字列の探索木
- 最初はRootのみ
- MCTSを用いて探索木を徐々に伸ばす
- 中間ノードは、SMILESのprefixに対応
- RNNを用いて完全なSMILESに補完する



➡ 有効なSMILES生成の確率が高い(40%超)
高速な生成が可能

RNN : 25万化合物で訓練

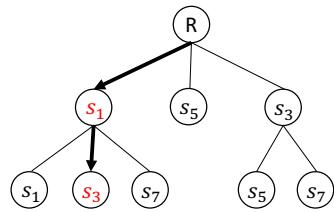
- 入力 : 文字列 s_1, \dots, s_T
- 出力 : 文字列の確率分布 $P(y_1), \dots, P(y_T)$
- 出力は、入力を右に一文字シフトしたもの



Keras & Tensorflow

MCTSによる探索

(a) Selection

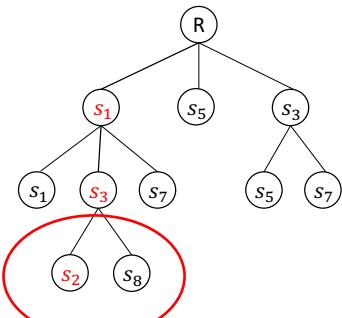


各ノードのスコア

$$u_i = \frac{w_i}{v_i} + C \frac{\sqrt{2 \ln V_{parent}}}{v_i}$$

子ノードの
スコアの平均
“伸びしろ”

(b) Expansion

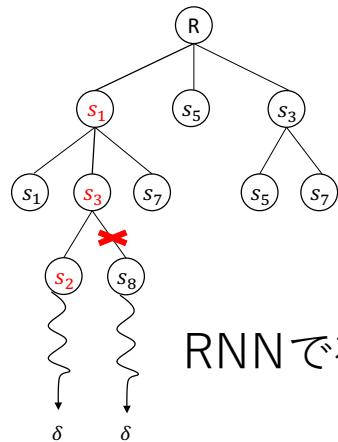


有効なSMILESのみ以下の
drug-likenessの値を評価

$$J(m) = logP(m) - SA(m) - RingPenalty(m)$$

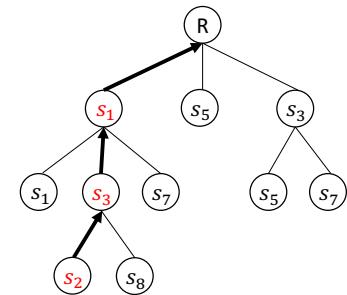
(Gómez-Bombarelli et al. ICML2016)

(c) Simulation



RNNで補完

(d) Backpropagation



全体のスコア
を更新

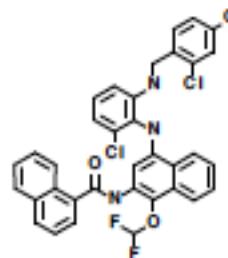
$$w_i = w_i + r$$

$$v_i = v_i + 1$$

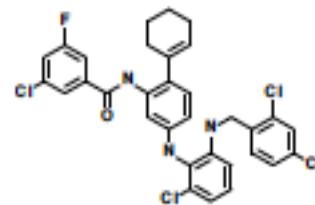
Table 1. Maximum score J at time points 2,4,6 and 8 hours achieved by different molecular generation methods. The rightmost column shows the number of generated molecules per minute. The average values and standard deviations over 10 trials are shown.

Method	2h	4h	6h	8h	Molecules/Min
ChemTS	4.91 ± 0.38	5.41 ± 0.51	5.49 ± 0.44	5.58 ± 0.50	40.89 ± 1.57
RNN+BO	3.54 ± 0.27	4.46 ± 0.24	4.46 ± 0.24	4.46 ± 0.24	8.33 ± 0.00
Only RNN	4.51 ± 0.27	4.62 ± 0.26	4.79 ± 0.25	4.79 ± 0.25	41.33 ± 1.42
CVAE+BO	-30.18 ± 26.91	-1.39 ± 2.24	-0.61 ± 1.08	-0.006 ± 0.92	0.14 ± 0.08
GVAE+BO	-4.34 ± 3.14	-1.29 ± 1.67	-0.17 ± 0.96	0.25 ± 1.31	1.38 ± 0.91

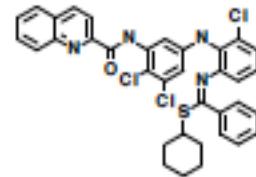
SMILES representation	<i>J</i>
O=C(Nc1cc(Ne2c(Cl)cccc2NCc2ccc(Cl)cc2Cl)c2cccc2c1OC(F)F)c1cccc2cccc12	6.56
O=C(Ne1cc(Ne2c(Cl)cccc2NCc2ccc(Cl)cc2Cl)ccc1C1=CCCCC1)c1ee(F)cc(Cl)c1	6.43
O=C(Ne1ee(Ne2e(Cl)cccc2N=C(SC2CCCCC2)c2eeee2)ee(Cl)c1Cl)c1eee2eeee2n1	6.34
O=C(Ne1ee(Oe2eee(Cl)cc2Cl)ccc1Ne1ee(Cl)ccc1Cl)c1eee(Cl)cc1	6.33
O=C(Ne1cc(Ne2c(Cl)cccc2Cl)c(Cl)cc1Br)N(c1cccc1)c1ccc(Cl)cc1	6.26
O=C(Ne1ee(Oe2e(Cl)cccc2Oc2ccc(-c3cccc3)cc2)ccc1Cl)c1eeee1	6.19
O=C(Ne1ee(Ne2e(Cl)cccc2Cl)c(Cl)e(C(=O)N(Ce2eeee2)c2eeee2)c1Cl)c1eeee1F	6.08
O=C(Ne1ee(Oe2eee(Cl)cc2Cl)cc(Cl)c1Cl)c1neoc1-c1eee(Sc2eeee2)cc1	6.007
O=C(Ne1ee(Ne2e(Cl)cccc2NCc2ccc(Cl)cc2Cl)c2ncccc2c1Cl)c1eee(Cl)cc1	6.0067
O=C(Ne1ee(Ne2c(Cl)cccc2NCc2ccc(Cl)cc2)c(Cl)cc1Cl)c1ee(F)cc1Cl	6.0062



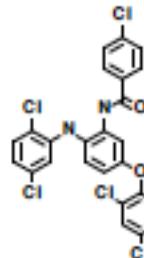
J = 6.56



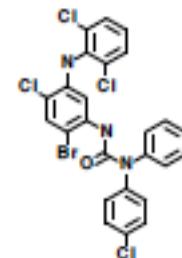
J = 6.43



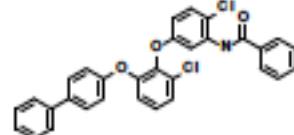
J = 6.34



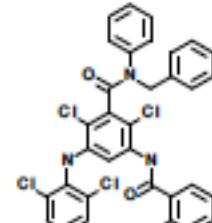
J = 6.33



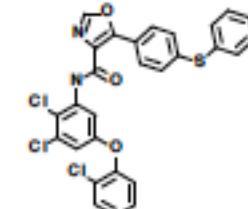
J = 6.26



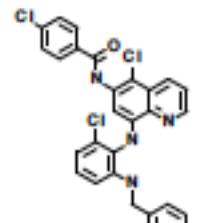
J = 6.19



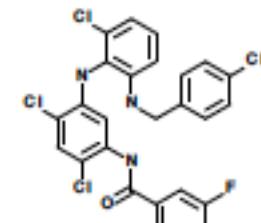
J = 6.08



J = 6.007



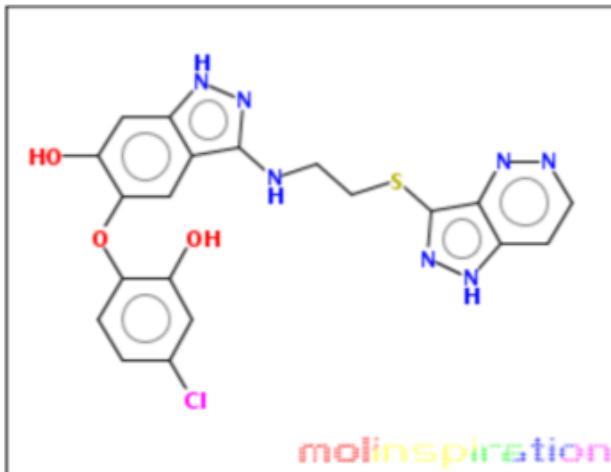
J = 6.0067



J = 6.0062

Docking to CDK2 (ChemTS + rDock)

originalSMILES Oc1cc(Cl)ccc1Oc1cc2c(NCCSc3n[nH]c4ccnnC34)n[nH]c2cc1O
miSMILES: Oc1cc(Cl)ccc1Oc1cc2c(NCCSc3n[nH]c4ccnnC34)n[nH]c2cc1O



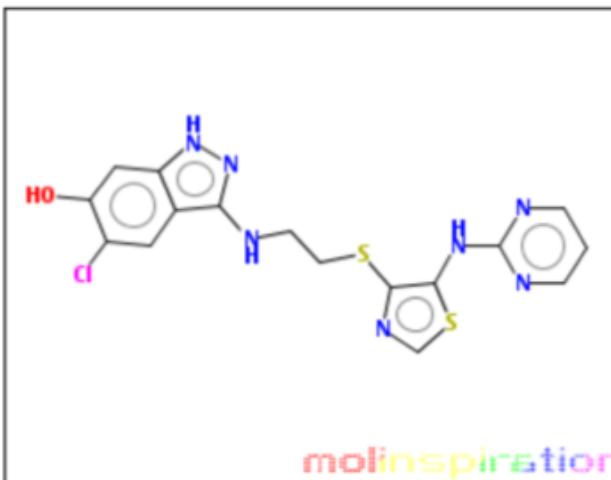
[Molinspiration bioactivity score](#) v2016.03

GPCR ligand	0.29
Ion channel modulator	0.13
Kinase inhibitor	0.64
Nuclear receptor ligand	-0.15
Protease inhibitor	0.05
Enzyme inhibitor	0.15

[Get data as text](#) (for copy / paste).

[Get 3D geometry](#) BETA

originalSMILES Oc1cc2[nH]nc(NCCSc3ncsc3Nc3nccn3)c2cc1Cl
miSMILES: Oc1cc2[nH]nc(NCCSc3ncsc3Nc3nccn3)c2cc1Cl



[Molinspiration bioactivity score](#) v2016.03

GPCR ligand	0.20
Ion channel modulator	-0.01
Kinase inhibitor	0.93
Nuclear receptor ligand	-0.43
Protease inhibitor	0.13
Enzyme inhibitor	0.19

[Get data as text](#) (for copy / paste).

[Get 3D geometry](#) BETA

おわりに

- Clever but slow (BO) 対 Stupid but fast (MCTS)
- Wanted: 合成ルート自動生成

The screenshot shows the Materials World magazine website. At the top, there's a dark header with the logo "mW MATERIALS WORLD" and a search bar. Below the header is a navigation menu with links: IOM3, HOME, ABOUT, BACK ISSUES, MEDIA INFORMATION, SUBSCRIBE, CONTACT, FEATURES, and NEWS. The main content area features a large image of a hand holding a smartphone displaying a game interface with a tree search visualization. To the left of the image, the title "GAME CHANGER" is displayed above a byline for "Ellis Davies". Below the title, there's a short summary of the article. To the right of the image, a vertical sidebar lists years from 2010 to 2017, each corresponding to a different news article.

GAME CHANGER

Ellis Davies
Materials World magazine, 1 Oct 2017

Ellis Davies reports on a method for designing advanced materials using an algorithm created to beat computer games.

An algorithm that identifies the best moves to beat computer games – the Monte Carlo tree search (MCTS) – has been used to develop a tool that allows researchers to determine the ideal placements for atoms within a structure to design advanced materials, such as metal and polymer matrix materials.

2017
2016
2015
2014
2013
2012
2011
2010

Monte Carlo tree search (MCTS) for a binary atom assignment problem. The