Can AI find “medicine-likeness”? Challenging Classification of Compounds (AI x Pharmaceuticals) [Paper]

- Commentary on medical papers
- masashi

Not all compounds can be drugs. Can AI find patterns in potential drug compounds?

Contents

: To streamline drug design

Theme: To predict potential drug compounds using machine learning

Purpose: To classify drugs and non-drugs

Challenge: I want to streamline drug design

Drug design is a method of synthesizing compounds that exhibit desirable physiological activities as drugs, and is one of the important techniques in the drug discovery process. It is expected that predicting the characteristics of compounds that make drugs effective will greatly improve the drug discovery process.

What research is actually being done on the subject of predicting the properties of compounds as drugs? I would like to introduce presentations by researchers such as Abraham Yosipof of the College of Law and Business in Israel.

They attempted to build a machine-learning model that classifies compounds into drugs and non-drugs by using features of their physicochemical parameters.

First, I summarized the mission, methods, and results of the research by Abraham Yosipof et al.

AI technology database

✔ **Predict**

the potential of mission compounds as drugs.

✔ **Solution method Using**

multiple machine learning models, compounds were classified into "drugs" and "non-drugs" based on physicochemical parameters.

✔**Results A**

classification rate of over 0.81 was obtained, and effective disease categories were also successfully classified.

I will start by explaining the mission.

Purpose: To classify drugs and non-drugs

In recent years, data mining and machine learning are attracting attention as technologies for efficient drug design. By exploiting these, it is believed that basic patterns of chemical and pharmacological features that are important in development can be found, and the design of compounds as drugs can be optimized according to those properties. Moreover, if we can classify compounds as drugs or non-drugs, we may be able to classify the diseases for which drugs are effective.

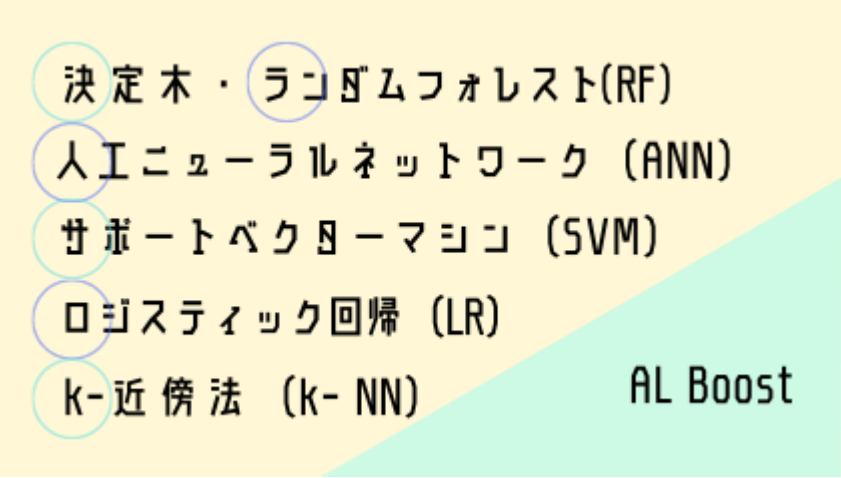
A research group of and colleagues built a machine learning model to classify compounds.

AI technology database

For the data set, 732 kinds of compounds were used from DrugBank , which is a compound database (drug: 366 compounds and non-drug: 396 compounds).

Physicochemical parameters were obtained using ChemAxon with 35 molecular features such as logP values and molecular surface areas.

Machine learning algorithms include decision trees , random forests (RF), support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbors (k-NN), logistic regression (LR), and A combined method, AL Boost, was used.



Result: Succeeded in classifying drugs and non-drugs with high accuracy

As a result, the classification rate (CCR) for AL Boost was good at 0.81, and the average CCR for the other six algorithms was 0.77.

J4.8	0.67	0.67	0.67	0.67	0.01	0.35	0.72	0.81	0.75	0.76	18.90	0.52
RF	0.76	0.76	0.76	0.76	0.00	0.52	0.80	0.86	0.82	0.82	10.21	0.65
k-NN	0.73	0.73	0.73	0.73	0.06	0.46	0.75	0.78	0.76	0.76	1.34	0.53
SVM	0.73	0.74	0.73	0.73	0.25	0.47	0.74	0.73	0.74	0.74	0.17	0.47
ANN	0.77	0.72	0.75	0.75	5.25	0.50	0.81	0.81	0.81	0.81	0.12	0.62
LR	0.76	0.76	0.76	0.76	0.01	0.52	0.77	0.70	0.74	0.74	12.58	0.48
AL Boost	0.76	0.77	0.76	0.76	0.01	0.53	0.80	0.81	0.81	0.81	0.22	0.62
Naïve Bayesian*	0.74	0.89	–	0.82	–	0.64	0.50	0.92	–	0.70	–	0.46

*Naïve Bayesian results were taken from García-Sosa and Maran (2013).

Classification of drugs and non-drugs

Furthermore, we predicted disease categories such as antitumor drugs, nervous system drugs, and cardiovascular drugs, and succeeded in separating them with high accuracy.

	Training set						Test set					
Model	Specificity	Sensitivity	CCR	Accuracy	Variance	MCC	Specificity	Sensitivity	CCR	Accuracy	Variance	MCC
J4.8	0.90	0.64	0.70	0.88	179.41	0.46	1.00	0.67	0.95	0.92	277.78	0.78
RF	0.93	0.82	0.77	0.91	28.81	0.63	0.96	1.00	0.88	0.96	4.73	0.85
k-NN	0.89	0.83	0.65	0.89	7.72	0.47	0.95	0.75	0.85	0.92	104.60	0.70
SVM	0.92	1.00	0.75	0.92	17.00	0.68	0.95	0.75	0.85	0.92	104.60	0.70
ANN	0.93	0.71	0.79	0.90	120.76	0.61	0.96	1.00	0.88	0.96	4.73	0.85
LR	0.92	0.89	0.74	0.91	1.93	0.63	0.96	1.00	0.88	0.96	4.73	0.85
AL Boost	0.92	0.89	0.74	0.91	1.93	0.63	0.96	1.00	0.88	0.96	4.73	0.85
Naïve Bayesian*	–	–	–	0.88	–	0.57	–	–	–	0.90	–	0.62

*Naïve Bayesian results were taken from García-Sosa and Maran (2013)

Classification of antitumor and nervous system drugs

From the above, it is expected that this model can be applied to predict the potential of compounds as drugs and the diseases for which drugs are effective.

That's all for the research introduction.

In the future, more efficient drug design for various diseases may be realized.

Check out other articles on " AI x Drug Discovery"

- How can the development of new antibacterial drugs be helped [Drug making in the age of AI]
- Can tuberculosis be defeated with drugs? It seems that analysis can be done with AI.
- Predict drug activity against cancer cells with AI? Other 5 latest research on drug discovery AI [Weekly]

***This article will be members-only content from tomorrow onwards. Similarly, all past articles can only be viewed by members. If you would like to view it, please register here ! (Member registration is free.)**

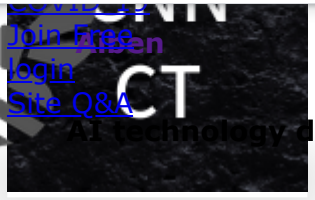
masashi Contributor's past articles

In graduate school, I studied medicine. I would like to delve into the use of AI mainly in the fields of drug discovery, manufacturing, and finance. Twitter: @masa05240112

Using AI to classify and predict fake news on the web (AI x web) [Paper]

Samsung develops AI technology to extend battery life of smartphones! Other 5 latest "AI patents" [Weekly]

Related article

1. 

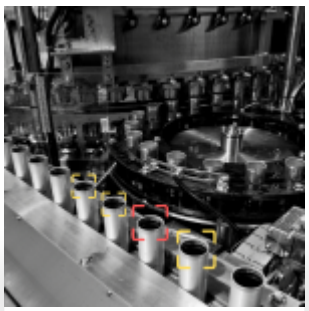
   Advances in materials development - Improving images with CNN! [AI paper]

2. 

   See the forefront of "Rakugo AI". Great research on whether machines can make people laugh (...

3. 

   High-level analysis of tourist preferences (China) [AI paper]

4. 

   Streamline the inspection process! AI technology to detect "defective steel cans" (AI...

AI technology database

5. 

[Is it possible to elucidate the mechanism of schizophrenia (AI x medical care) [paper]]

Recommended articles

1. 

["Check missing values in Pandas dataframe!" AI quiz]

2. 

["Make the array a pandas data frame!" AI quiz implementation version [...]

3. 

["What is the evaluation method after model creation?" AI quiz implementation version [Question 5]]

AI technology database

4. 

"Least-squares method with scikit-learn!" AI quiz implementation edition [Part ...

5. 

"Read csv files with pandas!" AI quiz implementation version [...

content search

search

Industry/Category

5G IoT Sponsored VR/AR web entertainment art sports sales marketing Security drone Mental health/psychology robot human resources Medical/Healthcare pharmaceuticals Politics and Society Education/Career Material patent surroundings self-driving manufacturing and design Agriculture and Food city/sightseeing Finance/Economy

New arrivals of original articles from SNS

Articles read in the last month

"Estimating the 'depth of an object' from a single image" A machine learning notebook committed to the field [vol.19]

List of "data analysis competitions" that can be entered from Japan [world version summary]

"I want to do useful research rather than business" Interview with attraction engineering researcher Toshihiko Yamazaki [Introduction to AI Researchers Vol.1]

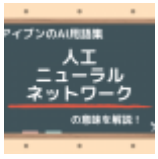
Released an AI model in which 3D humans move according to verbal instructions

What is a "classifier"? Quickly explain the meaning! [AI Glossary]

How advanced is AI power prediction? Research case summary

It seems that deep learning can understand the "pain" of experimental mice (AI x pharmaceuticals) [Paper]

Accuracy of age estimation never stops! Featured New Method (AI x Marketing) [Paper]



What is an "Artificial Neural Network (ANN)"? Quickly explain the meaning! [AI Glossary]

University of Tokyo AI Beginner Learning Diary Day 3 "Polynomial Curve Fitting and Overfitting"

Is it an era when lip reading can be done with AI? Be careful with your lips (AI x IoT) [Paper]

What is "Mean Absolute Percentage Error (MAPE)"? Quickly explain the meaning! [AI Glossary]

AI technology database

E qualification study content summary Day 1 “Overview of deep learning”

What is "MobileNet"? Quickly explain the meaning! [AI Glossary]

PAGE TOP

- HOME
- What is Aibn
- Aibn management member list
- Company website
- inquiry
- YouTube