COVID-19 Drug Discovery Using Genetic Constrained Graph Variational Autoencoder (GCGVAE)

Tianyue Cheng
Beijing No.4 High School International Campus
No. 2 Zongmao Toutiao, Xitiejiang Alley, Beijing
tianyue.cheng2002@gmail.com

Tianchi Fan
Beijing No.4 High School International Campus
No. 2 Zongmao Toutiao, Xitiejiang Alley, Beijing
fan_tianchi@163.com

Landi Wang
Beijing No.4 High School International Campus
No. 2 Zongmao Toutiao, Xitiejiang Alley, Beijing
zhzhuang22@163.com

August 20, 2020

Abstract

Variational Autoencoder is an effective and efficient tool while dealing with various generative problems. In this paper, we developed a unique kind of variational autoencoder, which is Genetic Constrained Graph Variational Autoencoder (CGCVAE), to discover the potential drug for Severe Acute Respiratory Syndrome Corona-Virus 2 (commonly known as SARS-CoV-2). After evaluating our results with tools that indicates the effectiveness while inhibiting SARS-CoV-2, we concluded that our results are much more effective than those existing drugs for treating COVID-19. Our model could also be used to generate other drugs given the assays of target protease.

1 Abbreviation

Table 1: Abbreviation List

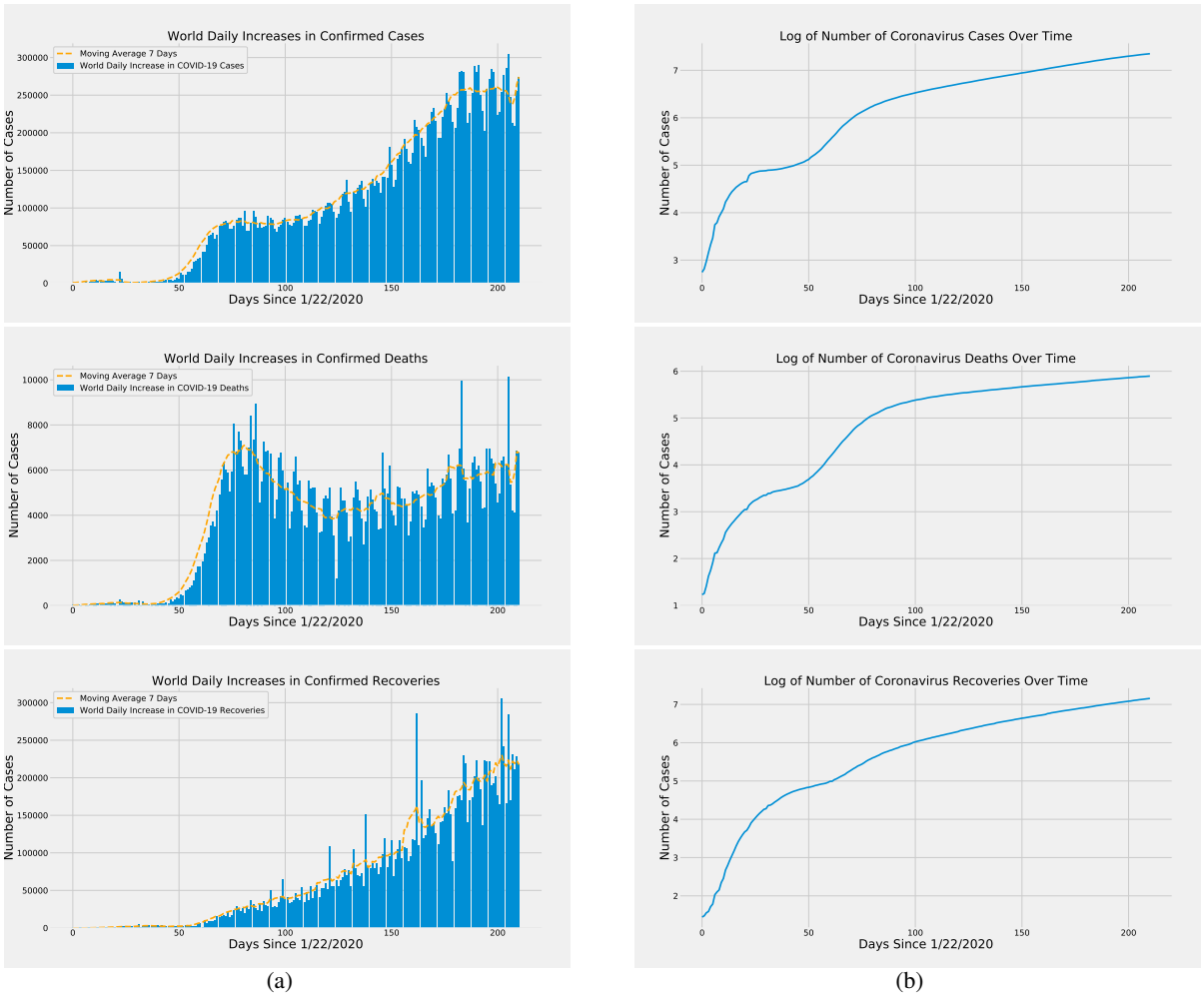
Full Name	Abbreviation
Severe Acute Respiratory Syndrome Corona-Virus 2	SARS-CoV-2
Coronavirus Disease 2019	COVID-19
Genetic Algorithm	GA
Machine Learning	ML
Deep Learning	DL
Edge Memory Neural Network	EMNN
Generative Adversarial Network	GAN
Simplified Molecular Input Line Entry System	SMILES
Variational Autoencoder	VAE
Constrained Graph Variational Autoencoder	CGVAE
Genetic Constrained Graph Variational Autoencoder	GCGVAE
Reinforcement Learning	RL

2 Introduction

2.1 Motivation

COVID-19 firstly been found at Wuhan, a city in China. The virus that caused COVID-19 is designated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was also previously known as 2019-nCoV. The most common symptoms of COVID-19 are fever, tiredness, and dry cough. Some patients even have aches and pains, nasal congestions, runny nose or sore throat. These symptoms are usually mild and begin gradually. Even though most people will recover from COVID-19, some still have serious illness, especially for people aged over 60. According to WHO (World Health Organization), around 1 in every 5 people who are infected with COVID-19 develop difficulty in breathing and require hospital care.

Although different countries carried out policies to stop the virus and cure the sufferers, many people are still exposed to sever threats. People over 60 often have underlying medical conditions such as diabetes, heart diseases, respiratory disease or hypertension, which exposes them at greater risk if they had been infected by the virus. In Indonesia, the 2018 Riskesdas basic health survey data show that cardio vascular disease and diabetes are among the highest disease burdens of the country. Nearly 11 % of Indonesian adults have high blood sugar levels and 1.5% suffer from heart disease – making these groups vulnerable to developing severe COVID-19 symptoms if they become exposed to the disease. Moreover, the same Riskesdas data show that nearly 63% of adult men in Indonesia smoke. This number is among the highest prevalence of smokers in the world.



(a) (b)

Figure 1: The statistics of COVID-19 since January

Thus, it is necessary to find ways to develop new strategies to stop the virus and cure the infected. Many countries had put a lot of efforts in developing vaccines to enable the majority to be immune to the virus. Even though vaccines

is effective for most developed nations, in most developing countries, new vaccines could be extremely difficult to develop. Since developing countries have more risk of suffering from COVID-19, it is necessary to design a drug that could cure the infected effectively. Since mass producing drugs for COVID-19 is much cost-effective than developing vaccines, our results is useful for countries that urgently need drugs to cure the coronavirus.

When coming to sorting out all drugs in current database, ML algorithms performs much more effective and efficient than humans. Since there are millions of ligands and protease in the National Center for Biotechnology Information (NCBI) database, it can be extremely difficult for humans to find and develop drugs for COVID-19. In this paper, we developed a generated model called GCGVAE that could automatically sort all useful data from the database and generate useful moleculCres to bind with SARS-CoV-2 protease. Since its performance (measured by binding affinity) is much better than those existing drugs, including Remdesivir, Ribavirin, Umifenovir, Favipiravir, Lopinavir, etc., our generated molecules have great potentials to save more infected people around the globe.

2.2 Challenges

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

2.3 Significance

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

2.4 Organization of Report

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

3 Related Work

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

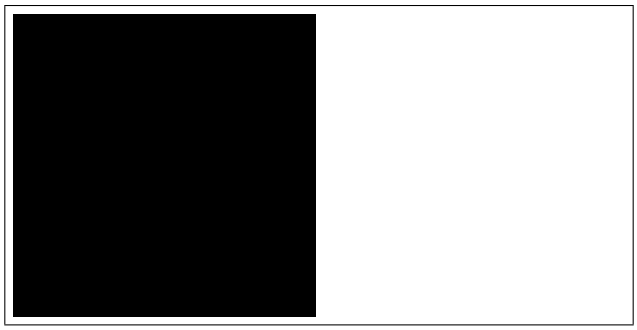


Figure 2: Sample figure caption.

test.png

4 Methods

4.1 Generative Model

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui. [1, 2] and see [3].

The documentation for `natbib` may be found at

4.2 EMNN and QED prediction

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi. See Figure 2. Here is how you add footnotes. [1] Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

4.3 GGNN model

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo. See awesome Table 2.

[1] Sample of the first footnote.

Table 2: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

4.4 Graph-Based Genetic Algorithm

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

References

[1] George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014.

[2] George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 312–318. IEEE, 2014.

[3] Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.