

<THE DISSERTATION TITLE>

By

Shuming Zhou

1611144

Supervised By

Nanlin Jin

A DISSERTATION

Submitted to

Xi’an Jiaotong-Liverpool University

in partial fulfillment of the requirements

for the degree of

MASTER OF RESEARCH

<THE DATE>

ABSTRACT

xxxx

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

Shuming Zhou

ACKNOWLEDGEMENTS

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

**TABLE OF CONTENT**

[LIST OF TABLES vi](#_Toc95215279)

[LIST OF FIGURES vii](#_Toc95215280)

[Chapter 1. Introduction 1](#_Toc95215281)

[Chapter 2. Literature Review 2](#_Toc95215282)

[Chapter 3. Research Methodology 3](#_Toc95215283)

[Chapter 4. Results 4](#_Toc95215284)

[Chapter 5. Analysis 5](#_Toc95215285)

[Chapter 6. Discussion 6](#_Toc95215286)

[Chapter 7. Conclusions 7](#_Toc95215287)

[REFRENCES CITED 8](#_Toc95215288)

[Appendix A. Xxxxx 9](#_Toc95215289)

[A.1 Xxxxxxx 9](#_Toc95215290)

LIST OF TABLES

Page

[Table 1 Principle Topics in Pattlets 11](#_Toc61607958)

[Table 2 Full Pattern-Description Organization 13](#_Toc61607959)

{To update the LIST OF TABLES, right click on the entries, chose “Update Field”, choose “Update entire table” and click on OK}

LIST OF FIGURES

Page

[Figure 1. Navigational Data-Entity Hierarchy (typical) 14](#_Toc120440498)

[Figure 2. Converse Matching of Association Types in Data-Entity Relationships 15](#_Toc120440499)

{To update the LIST OF FIGURESS, right click on the entries, chose “Update Field”, choose “Update entire table” and click on OK}

# Introduction

## Background

### Covid 19 situation

From 2020, COVID-19 has been a worldwide pandemic. There are many researches related to that topic. Each country takes their own policies and strategies to deal with this pandemic. Thus, it will be helpful to build a system which can help government to forecast the developing of COVID-19. According to WHO (2022), the current global infected population is 626,337,158. COVID-19 is mutated from SARS-CoV, it can bring different symptoms such as Tinnitus, gingivitis, sudden hearing loss, Bell’s palsy (Elibol).

COVID-19 also brings many social issues such as social distancing. Social distancing avoids people from offline meeting which will bring high social costs (Saladino, Algeri and Auriemma).

Thus, it is useful to predict the development of COVID-19. The prediction results can be used to formulate new policies to decrease the influence of COVID-19 to the public. Thus, predicting the COVID-19 is the focus of this dissertation.

### Brief introduction of previous research knowledge

目前的研究分为以下几个方向

对新冠疫情的类型分三大主体：对新冠疫情下的人民的研究(和上面相同)，对新冠疫情本体的研究，比如病毒本身的研究(https://www.sciencedirect.com/science/article/pii/S1567134820302537)医疗方面(https://www.sciencedirect.com/science/article/pii/S1521661620303181)以及对新冠疫情发展的预测(file:///Users/mac/Documents/GitHub/SAT405\_program/exercise/Prediction\_COVID\_19.pdf)，以及对政府的研究：政策的研究(https://www.nature.com/articles/s41562-020-0909-7)。（这三个）

对人民的研究主要是对疫情的心理和生理的评估，和本研究关系不大，而对病毒本身的研究以及医疗方面的研究涉及具体的医学和生物学知识，不在本文的探讨范围之内，其次就是对疫情的未来发展的趋势，这是对本文极其相关的点，具体也会在下文探讨。

对政府的研究主要是对NPI的研究，（https://www.sciencedirect.com/science/article/pii/S0048969721005982）NPI代表了具体的政策实施。当然也有对政策经济代价的研究，比如(https://www.medrxiv.org/content/10.1101/2020.03.26.20044552.abstract)，提到了中英两种模式下的经济代价。

再具体一些，对于这些研究，很多都是对单个因素的研究，比如单纯探讨疫情下人的心理生理状况，以及单独对npi的研究，其中也有很多是两个要素综合起来研究的，比如oxford的论文(file:///Users/mac/Documents/GitHub/SAT405\_program/405%E6%96%87%E7%AB%A0.pdf)，探讨病毒传播率以及NPI的政府政策，以及

<https://www.jmir.org/2020/9/e21419/>

这个文章主要是探讨政策和人民的态度。

当然最和本研究主题相关的，是政府政策以及病毒传染率这两个要素。对个人态度选择而言并不在本topic的论述范围之内。

There are several research directions towards COVID-19 prediction. The SIR with its derivatives (SEIR), the agent-based model, heuristic algorithm (genetic algorithm), machine learning and deep learning algorithm and game theory.

Below is the brief introduction of the models of these research directions.

#### SIR model

SIR represents different population in the COVID-19 pandemic.

S = S(t) = susceptible population

I = I(t) = infected population

R = R(t) = recovered population

This model uses these 3 parameters to construct equations and formulations to simulate the development of the COVID.

https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

#### Agent-based model

There are many agents in the model and each agent will have its states and actions. For example,

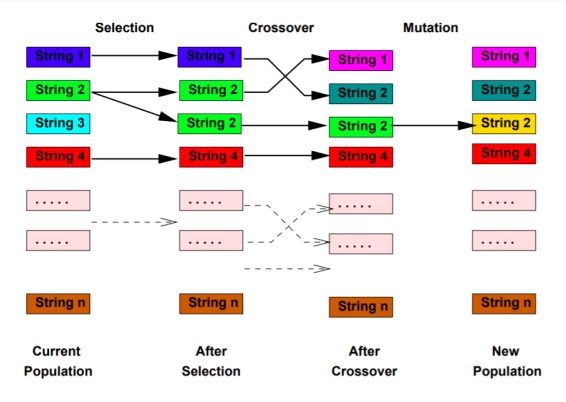
Shamil et al. (2021) defines each person as an agent and susceptible to COVID-19. Each agent will have 5 different states: HNASD to show if they are infected or healthy. Also, each agent will have their own profession and each profession will have corresponding tasks.

Agents are associated with groups based on their tasks and will interact with each other. There are 5 different groups to separate these agents: stay at home, commute, work or attend school, attend event stay at the hospital. In each group, the agents will be allocated to another group in order to realize the transmission in order to change their states.

#### Genetic algorithm

The genetic algorithm is used to let the population keep evolving and use fitness function to select the good one in order to make the population change over period. For example, according to the oxford paper, the “population” is the policies. The initial policies combination will evolve over time and it can then get the newly changed combinations of the policy.

GA (genetic algorithm) is a sort of algorithm which is used to simulate the natural selection process by Darwin according to Mathew (2012). Each individual will meet selection in natural environment. Those who does not adapt to the environment will eliminate while those fit the environment will survive. There are also mutation in the evolutionary process, the mutation will bring the offspring will traits which their predecessor do not have. Parents can crossover to bring their traits to their offspring. The GA has the characteristics described above: mutation, selection and inheritance. Each individual in GA will be represented as a genome. That is, a string of characters or numbers, just like the gene in the natural environment. Typically, the characters of the string are numbers, for example, each bit is 0 or 1. For example, if there is a need to represent the drink in the market, the drink can be accessed by 3 criteria: has sugar or not, has mineral substance or not, has fat or not. Therefore, the purified water can be represented as 0-1-0 which means it does not have any sugar but it has mineral substance and do not have any fat. A specific GA can be described like the flow chart below:



There are many turns in the entire GA loop process. Firstly, each string of characters will be initialized to get the current population, each string will be set as the combination of 0s and 1s. Secondly, selection will be made, those who does not fit will eliminate but those fit will survive. The GA has a function to judge if an individual adapts to the environment or not, that is the fitness function. The input of the function is the string of each individual. In figure 1, the string 3 is eliminated. After selection is the crossover process. String 1 and string 2 do such a process and generate the new offspring: new string 1 and new string 2. In order to describe the process of the crossover, it is very convenient to make an example: assuming that string 1 =1-0-0-0-0-1, string 2 =1-1-1-0-1-0. Then the new offspring will be new string = 1-0-0-0-1-0. The 1-0-0 at start is from the top 3 bit of string 1, 0-1-0 is from the bottom 3 bit of string 2. The last stage is mutation. In this stage, each string will have a chance to mutation. For example, if string =0-0-1-1-0-1, typically the mutation process will make each bit of the string to mutate. The third of the string is 1, then after mutation, the 1 will be changed to a value which is not the original value 1. Since here each bit can only be 0 or 1, thus it can only be 0 after mutation. Each bit will have a mutation probability to decide if that bit will change or not. After mutation, that loop is over and turn to the next loop.

#### Machine learning and Deep learning

Machine learning related to predict COVID-19 data are mainly supervised learning. To be specific, it is mainly the regression not the classification since the output data of the model is continuous values.

Deep learning is mainly about building the neural network, the input data will be fed into the model, each node of the neural network will use a function and the weight of the signal to do the calculating.

#### Game theory

The players of game theory is the most significant factors. Each player will have his/or strategies and each strategy will have a corresponding paid-off. The paid-off can be simply seen as the money got. It also have minus values and plus values which means get profit or pay for something.

## Research gap

Based on the previous introduction of the previous research and the forward literature review, there are lack of policy factors and real time data involved in simulating and predicting the COVID-19. However, the policy factor is a significant factors to be considered. For example, the re-opening economy strategy of UK will cause loosing the prevention of COVID-19 (International Monetary Fund, 2022). By comparing the new cases of UK and China and their anti-virus policies, a hypothesis can be made: the policy of a country will influence their own COVID-19 development (Worldometer, 2022). Thus, it is necessary to take policy into consideration.

## Topic Of Research Paper

Based on the brief introduction of previous research knowledge, the topic of this dissertation can be driven. This paper is mainly focusing on the policy factors in predicting the COVID-19 development, to be specific, its daily new cases. The question to be solved is how the policy will influence the prediction of Covid-19 daily cases. In other words, if the policy factor is introduced, will prediction of Covid-19 daily cases be more accurate or less accurate? How much accuracy or how much influence will that be? Thus, new models with policy factors will be introduced to answer this research question and its accuracy will be used to evaluate the answer.

# Literature Review

## Oxford Genetic algorithm by Vie (2020)

### Data

There is no real data used as input.

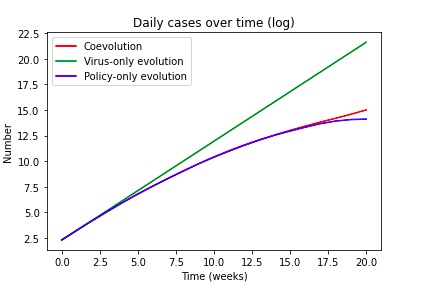
### Method

It uses genetic algorithm. By using the genetic algorithm, the changing of the policy combination can be simulated during the entire process.

### Results

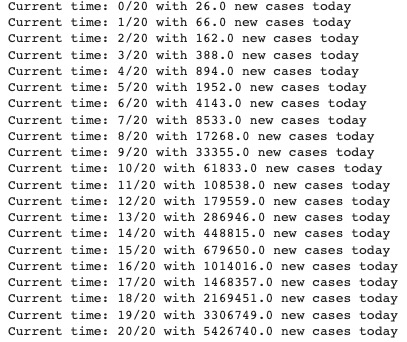
There are 3 different situations in the output.

1. Both the virus and the policy will evolve
2. Only virus evolve
3. Only policy evolve



### Gaps

It is obvious that there is no real time data. Therefore, when looking at its results and comparing to the real data, there is a huge difference between them.



This is the simulated data given by Oxford algorithm. It is obvious that in the simulated results, the new cases each time will only increase but not decrease. That is different from the reality.

The reason may be the parameter of this model is not set correctly. It said the parameter can be easily changed to get different result. However, it is still useful to use genetic algorithm as a simulating tool for qualitative research to predict the tendency of Covid-19.

## Machine learning and deep learning algorithm by Punn, Sonbhadra and Agarwal (2020)

### Data

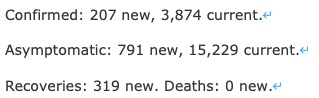
The data is achieved from the github website by Johns Hopkins whiting school of engineering (2022). The webpage claims its data is from Johns Hopkins University with different countries. For example, the attributes of data from UK daily cases are shown in the data below.

|  |
| --- |
| areaType |
| areaName |
| areaCode |
| Date |
| newCasesBySpecimenDate |
| cumCasesBySpecimenDate |
| new-FirstEpisodes-BySpecimenDate |
| cum-FirstEpisodes-BySpecimenDate |

It claims that it is valid since it is directly extracted from government website.

Actually, most of the data resource is government website so the data quality is good.

However, there are also some problems with this github website. Since this website is just a collection of these datasets, it does not do any reorganizing or improvement. Different countries have different table columns and are not unified, it may cause problems when doing unified data pre-processing. For example, the dataset of China has such fragment:



Which is different from the UK dataset.

And according to that paper, only 6 parameters of the dataset columns will be used: province/state, country/region, last update, confirmed, death and recovered cases.

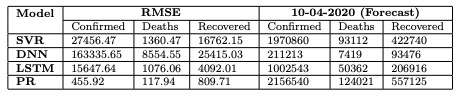
### Method

There are both deep learning and machine learning algorithms in that paper. These algorithms are implemented by python “sklearn” library, such as SVR, DNN, LSTM and PR.

### Evaluation

The mean squared error (MSE) is the most widely used objective function and root mean square error (RMSE) as a metric function for evaluating the regression models. [MSE 和RMSE的介绍]

### Results



### Gaps

The evaluation methods are not enough. For example, the R2 value is not included as the input. Also, the model does not include the parameters related to policy. In other words, policies corresponds to COVID-19 are not taken into consideration.

However, deep learning methods and machine learning methods are good to predict the new cases.

## SIR model by Wang et al. (2020)

### Data

Data is from China CDC.

### Method

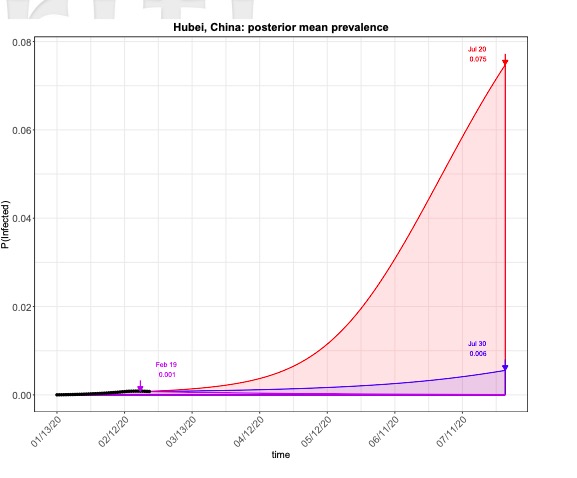
Using SIR model to predict the future data.

### Evaluation

However, it is now hard to evaluate the model. In that paper, the author put the results into the website to indicate his predicted data and real data there to see their differences. But, the website now is dropped and no data can be found there.

### Results

It uses a graph to show the result, it is actually only a vision since it provides 2 cases: the loosening case and strict case.



### Gaps

It is obvious that it uses many Math tools so it is based on solid theory. However, it is quite complicated because of using not only SIR but also Markov chain. Also, it does not pay attention to the policy factor.

## Agent-based model by Shamil et al. (2021)

### Data

2 categories of data are used: the Location-specific data and the Physiological data.

#### Location-specific data

1. demographics of the inhabitants in a particular city (i.e., education, employment, life expectancy, percentage of individuals having different professions, and the nature and timing of various tasks performed by the people) the data related to the number of transports and the average family size.
2. the data related to COVID-19 disease, its spread among the population, and the intervention measures taken by the authorities. These include the number of infections in the city and the day of the announcement of restrictive policies or awareness measures.

#### Physiological data

The probability of a person coughing and sneezing, touching contaminated objects, coming into physical contact with others, or washing hands is also an important parameter of our model, which would differ based on whether a person is at work, home, or hospitalized.

### Method

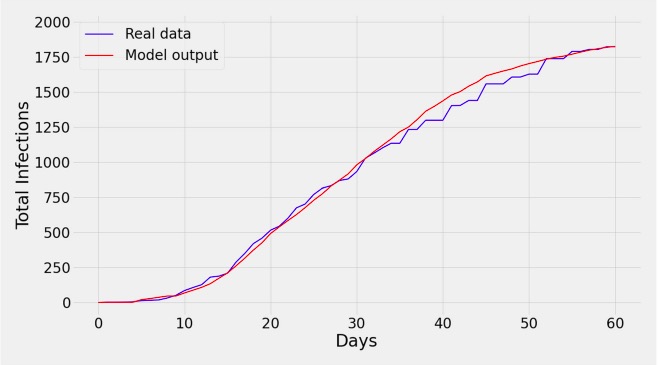
Since it is agent-based model, then the concept of agent should be explained. Each person in this model is an agent and susceptible to COVID-19. The states of each agent are 5 in total: healthy, asymptomatic, infected, symptomatic, dead or recovered.

Each agent is associated with a family and is assigned to 4 generic professions: healthcare workers, students, service holders and unemployed. Each profession has tasks represented as T.

### Evaluation

It uses its prediction data and the real data to make a plot to compare them in the time series

### Results



It is obvious that the predicted data and the real data is close to each other

### Gaps

The overall structure and the model and the data in use is concise and coherent and creative. However, in the result part, it only shows 60 days of data. The length of the daytime is not enough.

## Game theory by Kabir and Tanimoto

### Data

As the author wrote, the project has no additional data

### Method

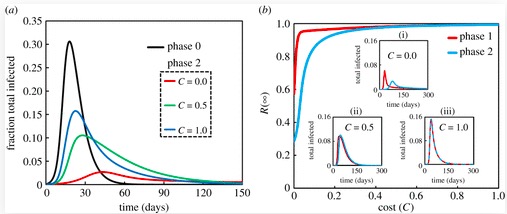
It uses the behavioural model which means each person in this model will have to choose 2 strategy: comply with stay-at-home or resist doing that. The stay-at-home order has some economic cost which is the classic attribute of game theory. Thus, by summing up the population, each strategy of the person can be calculated. Also, the SEIR model is also used to calculate the quarantined and infected population. It is then so-called SEQIHR(susceptible-exposed-quarantined-infected-hospitalized-recovered) model.

### Evaluation

Since there is no real time data, there is no evaluation methods involved, all it has is showing the result.

### Results

The result is about the predicted total infected population and the time series.



### Gaps

It is just like the genetic algorithm of Oxford, there is no comparison of the real time data. However, there is a good point of this model. In the complex parameters of this model, it mentions the paid-off of each strategy, the economic factor is included which is different from Oxford dissertation.

## Summary of these research

In summary, there are 2 different prediction directions: the accurate ones which can be used to predict the new infected population each day and the other one which just shows the tendency of the new infected population. In order words, one type is qualitative and another one is quantitative. The policy factors are mentioned on some of the research papers especially the Oxford one. However, the Oxford one does not pay attention to the economy costs of the policies.

In order to do the prediction, the dataset has to contains the policy data. According to the previous review, the Agent-based model is quite comprehensive thus making another agent-based model is not meaningful. The game theory model only pays attention to the personal paid-off of the strategy but the policy part is implicit: the 2 strategies of each person, compliance and non-compliance can be influenced by government. However, each person has their own values and decision, so their strategies can be influenced other factors. But the economic factors can be taken into consideration.

The SIR model doesn’t pay attention to policies, and it is so complicated.

The machine learning policy is coherent with evaluation methods and the testing and training models there with different ML and deep learning algorithms in python. However, it still does not have any policy factors.

The Oxford genetic algorithm pays attention mainly to policy outcomes and its infection population prediction is deviated from real time data. Also, it is better to invoke economic factors to this algorithm to make it more validate.

# Research Methodology

According to the literature review, there will be 2 different ways of method, the first one is qualitative research by simulation while another is quantitative research by calculating. These 2 will be both described below.

## Qualitative research

Qualitative research is to predict the tendency of the COVID-19 pandemic to see if the model can explain the development of COVID-19.

### Data

Because it is qualitative research, the main focus will be checking the tendency of the simulating development corresponds to the tendency of real data. Therefore, no additional data will be used or processed.

### Simulating algorithm

The algorithm used is adapted from genetic algorithm (GA), the brief introduction of GA can be found in 1.1.2.3. It is suitable to use GA since it can describe several objects and their inner attributes. Also, it can simulate the transformation of the status of each object, this will be expressed in detail in the following sections. In addition, GA is much comprehensive and simple than other simulation algorithms such as SIR and Agent-based algorithm.

There are several significant parameters and functions of this algorithm.

|  |  |
| --- | --- |
| *Attributes/Function* | *Explanation of the attribute/function* |
| Daily new cases infected population | The new population get infected each day |
| The policy | The policy chosen by the government on that day |
| Economy baseline | Each policy will bring its economic cost, government will balance the cost and infection |
| Policy change | The current policy will change according to different factors, this is represented as fitness function |
| Isolated population | Each day some people will be isolated |

这里写一个算法的伪代码

### Evaluation method

The evaluation method will use the tendency graph of adapted GA and the real data tendency graph. If the tendency of the real data tendency and result tendency graph are similar or the result can be used to explain the real data tendency, the reliability of the GA can be promised so that the influence of policy elements can be proved in a certain extent.

## Quantitative research

### data

#### data of the policy table

The data of the policy table is mainly from Coronanet Project (2023). However, there are so many columns (attributes) in this table and many of them are not related to the research, the detail data can be seen in Appendix. Here, the columns needed are given below.

|  |  |
| --- | --- |
| *Attribute* | *Meaning of the attribute* |
| type | The name of the policy |
| |  | | --- | | date\_start | | The start date of the policy |
| |  | | --- | | date\_end | | The end date of the policy |

By looking at the type attribute, all the policies are given below.

|  |
| --- |
| New Task Force, Bureau or Administrative Configuration |
| Anti-Disinformation Measures |
| Closure and Regulation of Schools |
| COVID-19 Vaccines |
| Curfew |
| Declaration of Emergency |
| External Border Restrictions |
| Health Monitoring |
| Health Resources |
| Health Testing |
| Hygiene |
| Internal Border Restrictions |
| Lockdown |
| Other Policy Not Listed Above |
| Public Awareness Measures |
| Quarantine |
| Restriction and Regulation of Businesses |
| Restriction and Regulation of Government Services |
| Restrictions of Mass Gatherings |
| Social Distancing |

#### daily cases data

The data is mainly from Our World in Data (2023). However, there are also many columns which should be eliminated in order to keep those related to the research.

|  |  |
| --- | --- |
| *Attribute* | *Meaning of the attribute* |
| date | Current date |
| new\_cases | New cases of infection on the day |
| total\_deaths | Total death population |
| new\_deaths | New death population |
| reproduction\_rate | R0 value of the virus |

### data pre-processing

The 2 tables of policy and cases should be combined in order to build a new table which has both policy influence and daily new cases in it. By taking this measure, with the same date, the current cases and the current policy combinations will be given as the input of the machine learning model. Below is the specific combining process.

1. All the concrete policies will be extracted from the policy table as columns to add into the cases.
2. Using the start date and end date of each policy to set the corresponding policies columns of that date into 1, it means on this date, government took these policies.

### Brief introduction of ML algorithm being used

#### Why Regression

Firstly, the target of this research is using many columns such as daily dead or taking a specific policy or not to predict just one column: the daily new case today.

Also, the label of the predicted column is given. In other words, the value of today’s new daily cases is given. Thus, the model has to be supervised learning instead of un-supervised learning since the value is predefined. Also, the value of new cases each day is not a discrete value, it’s continuous value. Therefore, it should not be a classification model but needs to be a regression model.

#### Regression

Regression is actually a type of question which has a form of Regression. Regression algorithm in machine learning is the combination of several different specific regression model such as linear regression.

This research will choose linear regression. There are different regression such as linear regression, Lasso regression and ridge regression.

Since the research question is not to explore which regression algorithm is most accurate and good for predicting, it is actually testing if the policy parameter has influence on COVID-19 development.

Thus, normal regression model and ridge regression are used and the result can be seen in Chapter 4.

#### Linear Regression

In linear Regression, there will be both independent and dependent variables. Independent variable is represented as xi (i=1,2,3) while dependent variable is represented as y (Hoerl and Kennard, 1970).

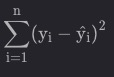
The function of the relationship between xi and y is

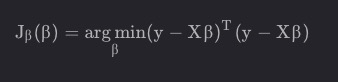


While 一部徐龙 represents the random error.

By feeding the data into the model, both Y and Xi will be given so that the coefficient like 贝塔0 will be adjust to get Y to be more close to actual Y^. Y^ is the original value while Y is predicted value.

In order to train model efficiently, it uses RSS to represents the differences between Y and Y^.

RSS = 

The formulation can be transformed into 

Solve the equation, then .

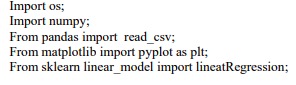
In order to get 贝塔, XTX has to be 满秩矩阵. However, this requirements can not always be satisfied. For example, if the tasks has so many columns which is even more than the row number, then XTX will not be a 满秩矩阵, then XTX 的行列式will be close to 0, so there will be more than one 贝塔satisfy this equation.

#### Ridge Regression

In order to solve the problem mentioned in the previous one, then Ridge regression is involved, by adding a lamda into the XTX, then the 贝塔can be get. In order words, Ridge regression is an variant of the ordinary linear regression

#### The brief introduction of sklearn library

Sklearn is a python library built for machine learning algorithms. It provides the template for Ridge regression and ordinary linear regression.

In order to use them, 

The importing part should be included.

Detailed information for sklearn for these 2 regression can be found in the official webpage (scikit-learn, 2023).

The regression approaches are trained, tested and used for prediction on real data using the mentioned column.

### Evaluation Method

There are many evaluation values of these 2 regression algorithms such as R2 and RMSE. These values should be used to evaluate the result of the algorithm and be introduced below. Also, these values of the results will be compared with results in past researches.

#### R2

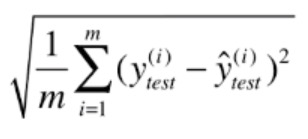
R2 = ESS/TSS.

Where ESS = 

And TSS = 

If R2 is close to 1, then the prediction is good if R2 is close to 0 (Shen and Zhang).

#### RMSE

RMSE = 

According to Davide, Matthijs and Giuseppe, if RMSE is close to 0, then the model is good, if it gets to 无限大, then it is bad.

The real dataset will be separated into training set and testing set. The mean\_squared\_error and R2 will be introduced to judge the accuracy of the model. Also, the plots will be generated as a result.

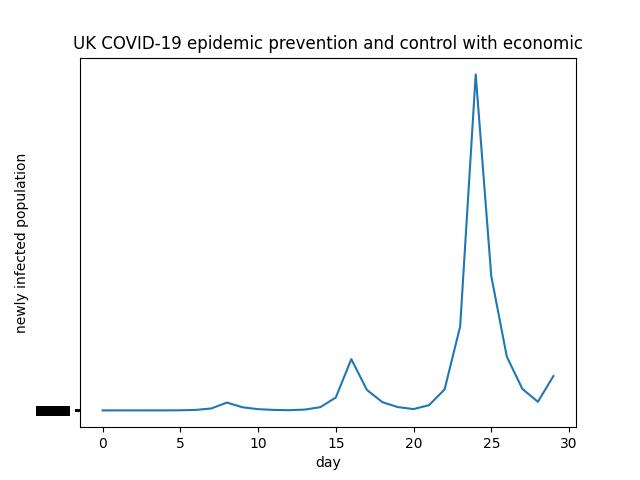
The python code can be found in Appendix.

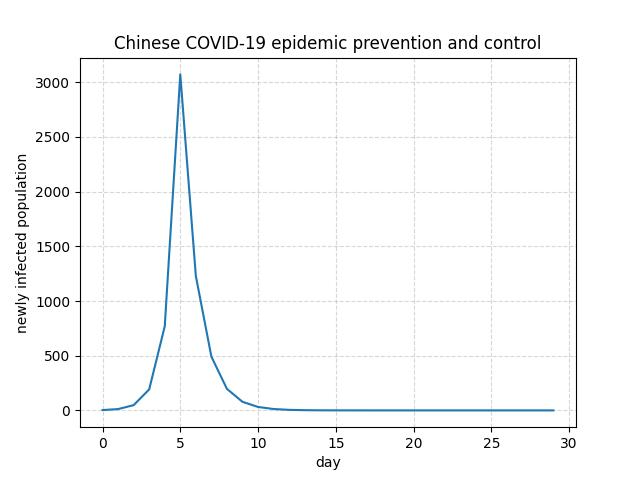
# Results

Qualitative research

The simulating results are shown below.

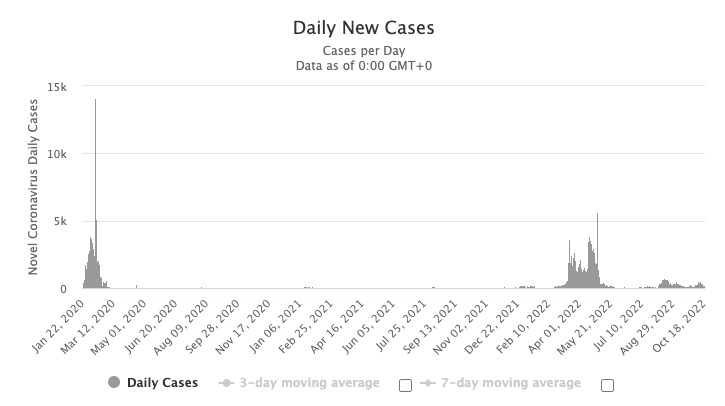
不能叫day，叫时期，需要解释一下



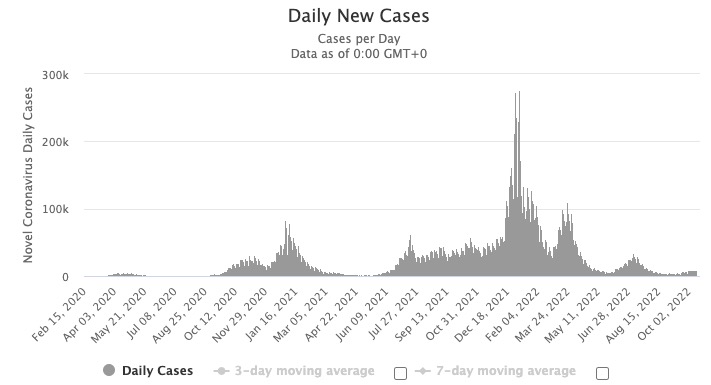


Below are the real data.

Chinese

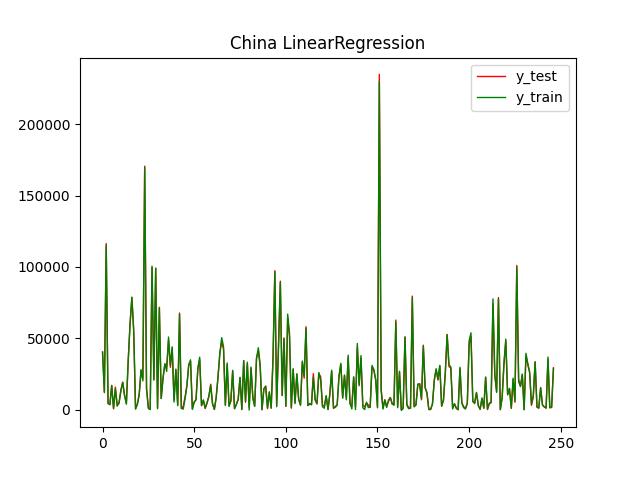


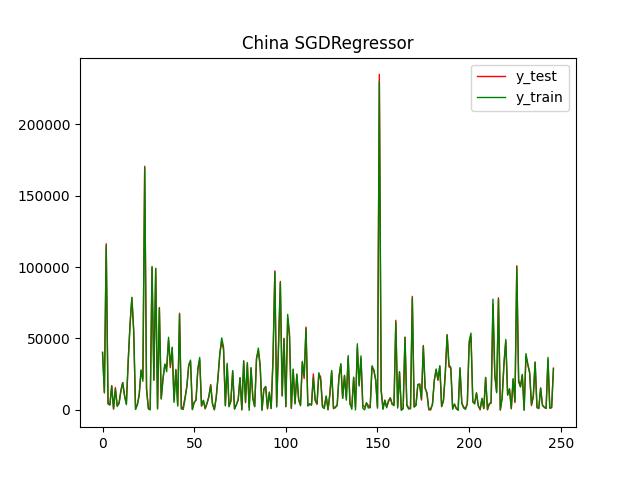
UK

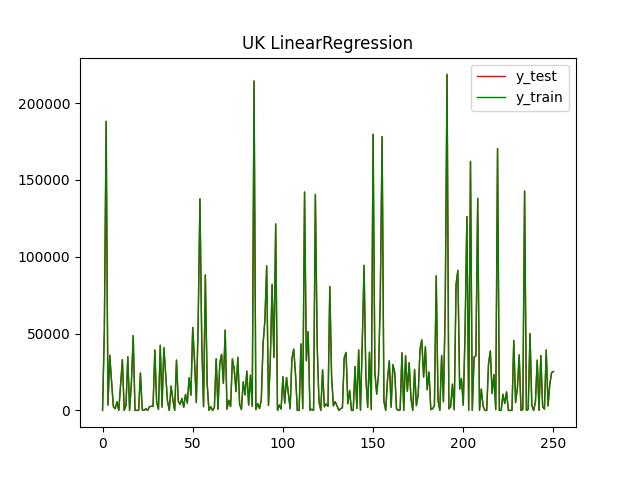


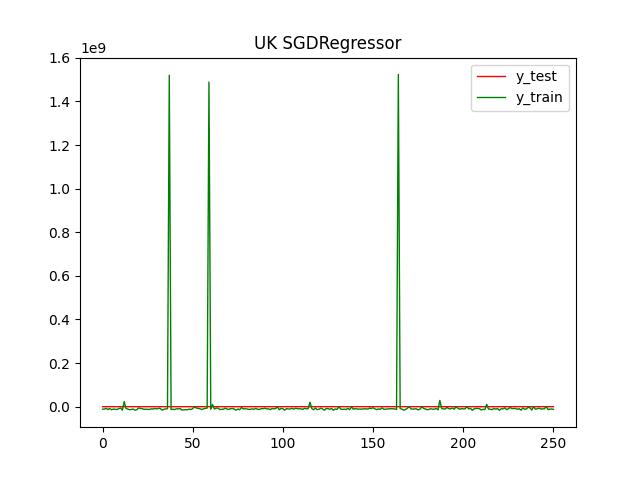
Quantatitive research

Both the training and testing data are shown below.









It is obvious that the SGDRegressor is much worse than LinearRegression. Other results parameters are given below.

UK

|  |  |
| --- | --- |
| LinearRegression |  |
| mean\_squared\_error | 6.956805139013924e-20 |
| r2\_score | 1.0 |

|  |  |
| --- | --- |
| SGDRegressor |  |
| mean\_squared\_error | 25739583591725.77 |
| r2\_score | -2013312.1793092396 |

Chinese

|  |  |
| --- | --- |
| LinearRegression |  |
| mean\_squared\_error | 4.22123953376573e-24 |
| r2\_score | 1.0 |

|  |  |
| --- | --- |
| SGDRegressor |  |
| mean\_squared\_error | 7097213485448.217 |
| r2\_score | -555132.8231911982 |

# Analysis

Xxxxx

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

# Discussion

Xxxxx

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

# Conclusions

Xxxxx

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

REFRENCES CITED

[] WHO (2022) WHO COVID-19 dashboard. Available at: <https://covid19.who.int/> (Accessed: 27 Oct 2022).

[] Elibol, E. Otolaryngological symptoms in COVID-19. Eur Arch Otorhinolaryngol 278, 1233–1236 (2021). https://doi.org/10.1007/s00405-020-06319-7

[] Saladino, V., Algeri, D. and Auriemma, V., 2020. The psychological and social impact of Covid-19: new perspectives of well-being. Frontiers in psychology, p.2550.

[] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67. Available at: https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634?tab=permissions&scroll=top

[] International Monetary Fund (2022) policy responses to COVID-19. Available at: <https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#C>

[] Rong, S., & Bao-Wen, Z. (2018). The research of regression model in machine learning field. In MATEC Web of Conferences (Vol. 176, p. 01033). EDP Sciences.

[] Our World in Data (2023) Coronavirus Pandemic (COVID-19). Available at: https://ourworldindata.org/coronavirus

[] Coronanet Project (2023) Coronanet Research Project. Available at: <https://www.coronanet-project.org/>

[] Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation." PeerJ Computer Science 7 (2021): e623.

[] scikit-learn (2023) sklearn.linear\_model.Ridge. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.Ridge.html#sklearn.linear\_model.Ridge

[] Worldometer (2022) COVID-19 CORONAVIRUS PANDEMIC. Available at: <https://www.worldometers.info/coronavirus/#countries>

[] Shamil, M., Farheen, F., Ibtehaz, N., Khan, I.M. and Rahman, M.S., 2021. An agent-based modeling of COVID-19: validation, analysis, and recommendations. Cognitive Computation, pp.1-12.

[] Mathew, T. V. (2012). Genetic algorithm. Report submitted at IIT Bombay.

[] Johns Hopkins Whiting school of Engineering (2022) COVID-19. Available at: <https://github.com/CSSEGISandData/COVID-19>.

[] Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., Kleinsasser, M., Barker, D., Eisenberg, M.C. and Song, P.X., 2020. An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China. Journal of Data Science, 18(3), pp.409-432.

[] Vie, A., 2021. Modelling SARS-CoV-2 coevolution with genetic algorithms. arXiv preprint arXiv:2102.12365.

[] Punn, N.S., Sonbhadra, S.K. and Agarwal, S., 2020. COVID-19 epidemic analysis using machine learning and deep learning algorithms. MedRxiv.

[] Kabir, K.A. and Tanimoto, J., 2020. Evolutionary game theory modelling to represent the behavioural dynamics of economic shutdowns and shield immunity in the COVID-19 pandemic. Royal Society open science, 7(9), p.201095.

[] Shamil, M., Farheen, F., Ibtehaz, N., Khan, I.M. and Rahman, M.S., 2021. An agent-based modeling of COVID-19: validation, analysis, and recommendations. Cognitive Computation, pp.1-12.

APPENDICES

###### Xxxxx

Xxxxxxx

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.