Covid-19 Measures Impact under Parametric Uncertainty

Filotas Theodosiou
a19filth
Data Driven Decision Making & Complex Data

Aim - Motivation

★ Unanswered Questions Regarding Covid-19
  - Conflicting Evidence on Studies
    - Survey Biases
    - Sampling/Measurements Errors

- The **World Health Organization** (WHO) reported an incubation period for COVID-19 between **2 and 10 days**. [1]
- China's **National Health Commission (NHC)** had initially estimated an incubation period from **10 to 14 days** [2].
- The United States' **CDC** estimates the incubation period for COVID-19 to be between **2 and 14 days** [3].
- DXY.cn, a leading Chinese online community for physicians and health care professionals, is reporting an **incubation period of "3 to 7 days, up to 14 days".**

Extra pressure on policy makers to turn uncertain knowledge into measures to tackle the pandemic.

★ How Policy maker should act?
  - What measures should we take based on the parametric uncertainty?

Main Idea

A Data Driven Decision Making Framework including

1. Visualization of uncertainty in scientific research
2. Possible Measures to tackle the spread
3. Forecast Model based on uncertain parameters

Pick the best combination of measures while considering the parametric uncertainty

A new Forecast based on user’s preferences will be produced

Target Users : Policy Makers

Experimental Design

★ Data Used
- 250.000 full-text articles regarding Covid-19
- Time Series data including cases, hospitalizations , deaths
- Extra Data Regarding Total Number of ICU beds

★ Models Used
- A Knowledge Extraction Model
- A forecasting Model

★ Forecast Evaluation:
- Mean Squared Log Error
- Mean Absolute Percentage Error

Knowledge Extraction Model

Analyze 250.000 Papers Regarding Covid - 19

Answer specific user queries with scientific findings

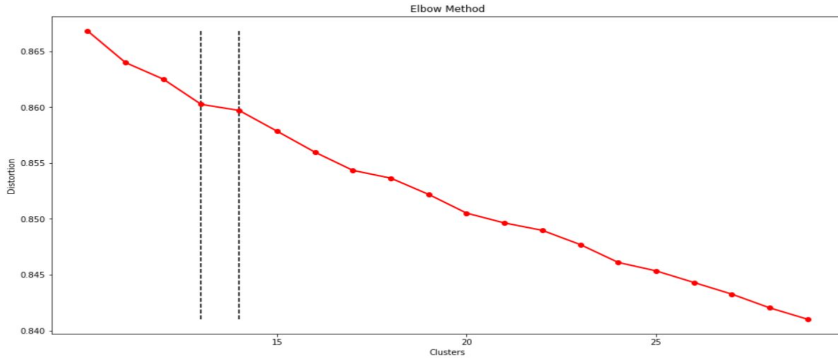
- What's the mean incubation period
- What's the proportion of asymptomatics

Queries are related to the epidemiological models parameters.

Many rephrases per query

Knowledge Extraction Model

- Reduce the volume of the corpus
    - Remove Non - English Papers
    - Keep Papers written on 2020
    - Relative Word Included on Abstract or Title
- Pre - Process
    - Assumption : Relative Papers After April
    - Remove Stop Words
    - PCA : Remove sparsity noise by reducing to 90%
- Cluster Documents
    - K - Means
    - Elbow method for optimal k



Cluster Selection Techniques

1. LDA topic modelling
    a. Extract Topics of each Cluster
    b. Get Keywords for each topic
    c. Combine Keywords for every cluster
    d. Return a cluster if the given query is included on its keyword list
2. Tf-IDF Similarity Measure
    a. Calculate Tf-IDF of every term on each document with a cluster
    b. Sum the scores for every term on the cluster
    c. Return the cluster with the highest score for every query
3. Combined Approach
    a. Return clusters found by each method
    b. Hides limitation of both approaches

Limitations :

- Many rephrases until a match is made

Advantages:

- Strong connection between query and cluster if match is achieved

Advantages:

- For every query and every rephrase, always return a cluster

Limitations:

- Connection between cluster and query might not be strong

Knowledge extraction algorithm

On the Remaining Papers

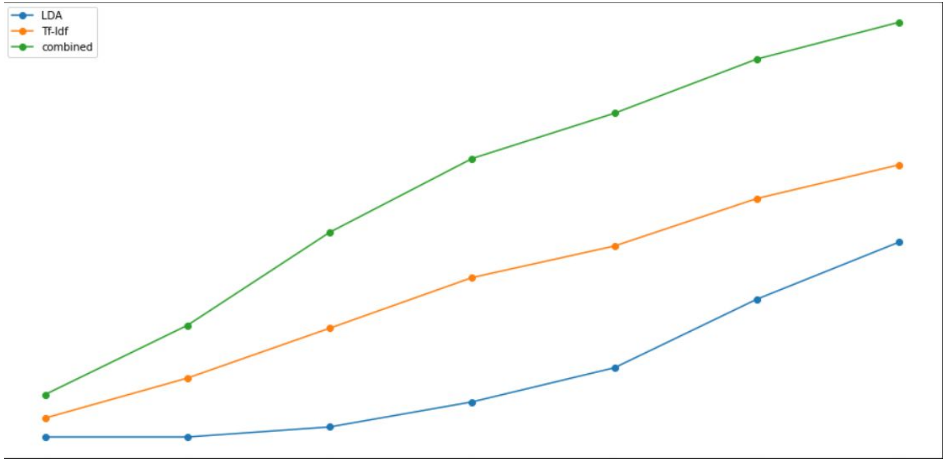
- Query word must be included in the abstract
- All Sentences which include a more detailed query are returned

Recommended Sentences are empirically evaluated.

- The value is added on the value list for each feature
- Several rephrases for each query are necessary for a good enough sample

Experiment Results: Quantity

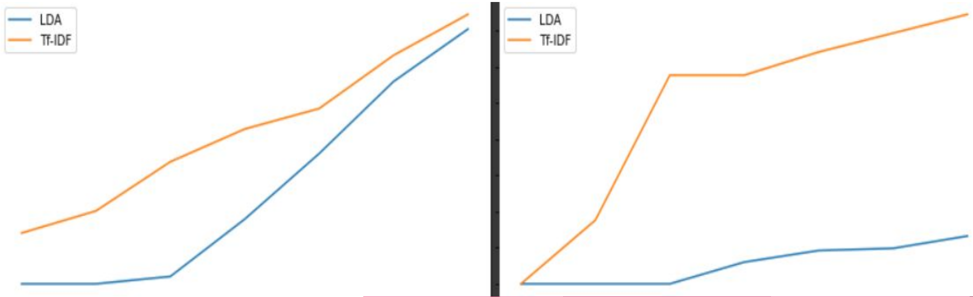
Mean cumulative sums of recommended sentences per different rephrase.



- Steady increase on returned results of Tf-IDF
- LDA requires more rephrases to find a good keyword-query match
- Combined Approach: Most returned Values

Combination of both methods:

- Best overall results
- If keyword-query match is not made -> Tf-IDF will be considered
- Returns results from both methods



Knowledge extraction algorithm

```
general = 'incubation'
detailed = 'incubation period days'
zz = knowledge_extraction_both(general,detailed,to_print=True)
```

```
Next Paper: incubation period and serial interval of covid-19 in a chain of infections in bahia blanca (argentina)
____
The estimated median incubation period in this study was 5.8 days for general transmissions; the estimated mean incubation wa
s 6.9 days which is 33% longer than the previously frequently adopted value-5.2 mean days as reported by Li (16)


The random-effects meta-analysis using restricted maximum likelihood (REML) was used to summarize the median incubation perio
d (days) and the corresponding 95% confidence interval (95%CI)



____
Next Paper: a systematic review and meta-analysis reveals long and dispersive incubation period of covid-19
____
The median incubation period 46 of COVID-19 is estimated as 5 to 6 days (1-4), while that of influenza A and B and SARS-CoV-1
are 1.4, 0.6 47 and 4.0 days, respectively (5)



____
Next Paper: estimation of the incubation period of covid-19 using viral load data
____
```

Feature Selection

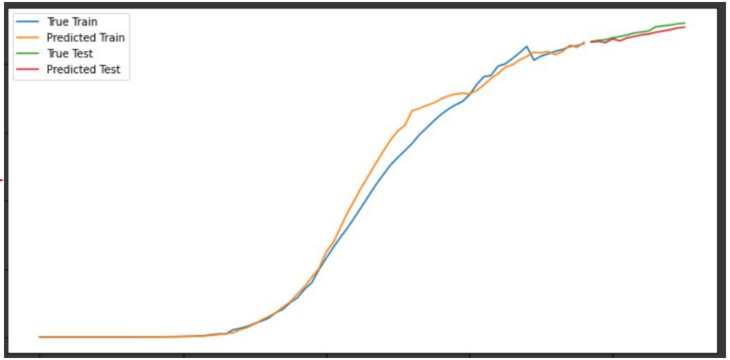
After experimenting with many combinations of features the following were picked and their values were extracted

The selection is related to the parameters of the forecasting model

1. Incubation days
2. Infectious days
3. Hospital Days
4. Days in Critical Condition
5. Mild - Asymptomatic Probability
6. Probability of Developing Severe Symptoms
7. Mortality After Developing Severe Symptoms

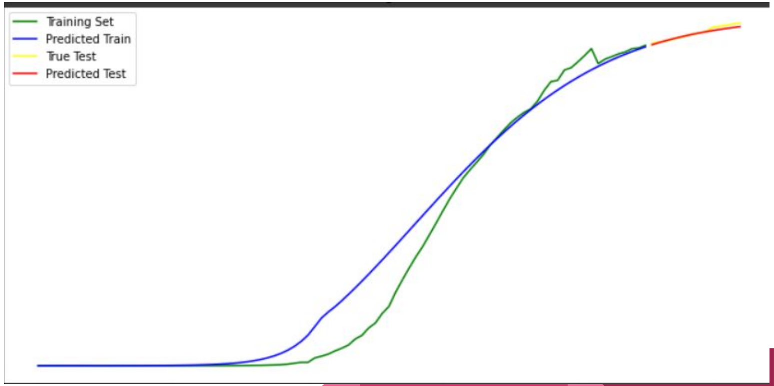
Forecasting Models

Best LSTM



Two models were compared:

Selected Model



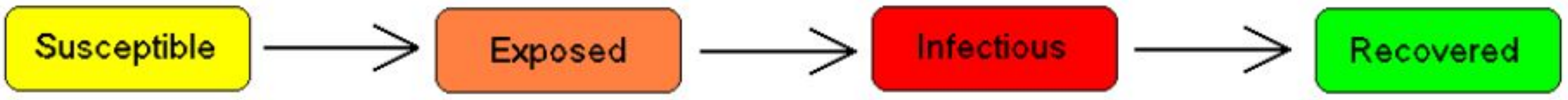
1. LSTM(w. various Variations)
    a. Bi-LSTM
    b. Single Layer LSTM
    c. Double Layer LSTM
    d. Ensemble Nets
2. Epidemiological Mathematical Models

Epidemiological Models

Forecast Number of Cases of a Given Country

- Model and Simulate Pandemics
- Transparent
- Highlight Feature Importance

Epidemiological Models



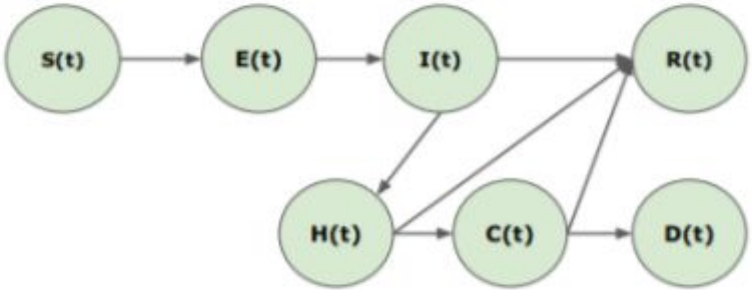
Split Population into epidemic stages

- Rate of Change - System of Differential Equations
  - Independant Variables reflect core dynamics
- Total Number of Differential Equations.
  - Trade-off : Complexity + Extra Factors
- Calibrate on True Data
- Initial Conditions

SEIR - HCD

Susceptible, Exposed, Infected, Recovered , Hospitalized , Critical, Deceased

Name	Description	Differential Equation	Explanation of Features
Susceptible	Population not immune to the virus	$\frac{dS}{dt} = -\frac{R_t}{t_{inf}} \cdot IS$	Rt-> reproduction number t_inf -> infectious period
Exposed	Population currently in incubation	$\frac{dE}{dt} = \frac{R_t}{t_{inf}} \cdot IS - \frac{1}{t_{inc}} \cdot E$	t_inc -> incubation period
Infectious	Number of infections actively circulating	$\frac{dI}{dt} = \frac{1}{t_{inc}} \cdot E - \frac{1}{t_{inf}} \cdot I$	
Recovered	Population no longer infectious due to recovery	$\frac{dR}{dt} = \frac{m}{t_{inf}} \cdot I + \frac{1-c}{t_{hosp}} \cdot H$	m -> prob of having mild or no symptoms t_hosp -> days in hospital before recovering or moving to ICU
Hospitalized	People who developed severe symptoms and moved to hospital	$\frac{dH}{dt} = \frac{1-m}{t_{inf}} \cdot I + \frac{1-f}{t_{crit}} \cdot C - \frac{1}{t_{hosp}} \cdot H$	f -> prob of passing out t_crit -> days in ICU(critical condition)
Critical	Hospitalized people whose condition got critical and moved to ICU	$\frac{dC}{dt} = \frac{c}{t_{hosp}} \cdot H - \frac{1}{t_{crit}} \cdot C$	c -> fraction of severe cases that turn critical
Deceased	People who passed out after being in critical condition	$\frac{dD}{dt} = \frac{f}{t_{crit}} \cdot C$	



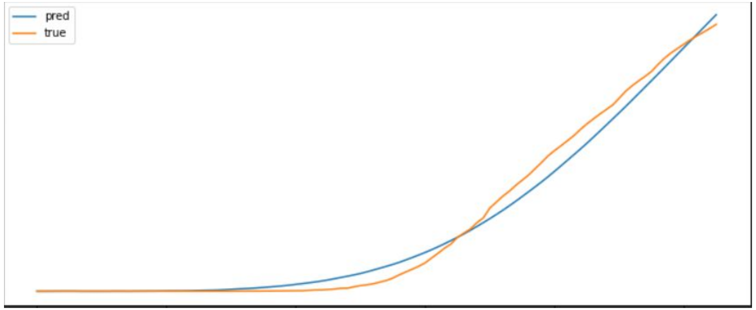
SEIR - HCD: Calibration

Equations' Independent Variables Selection:

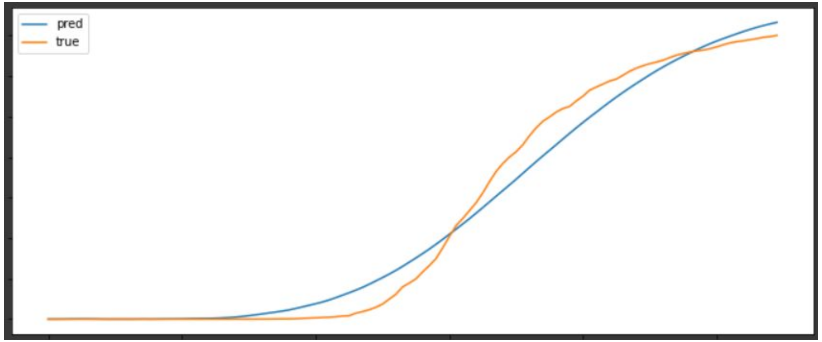
1. Optimize by calibrating on Real Data
   a. Fitting on real data
   b. Minimizing a cost function
2. User Selection

SEIR - HCD Examples

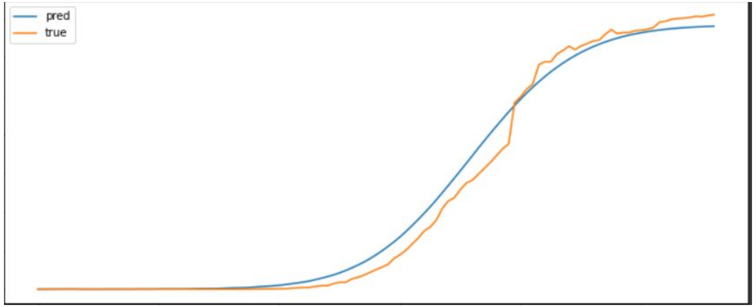
France



Italy



Germany



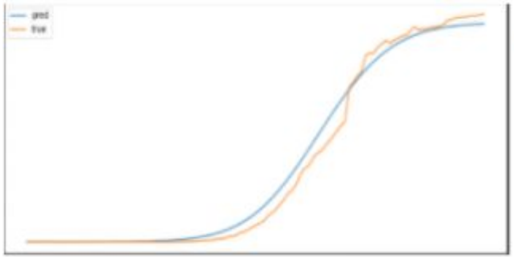
Forecasting Evaluation

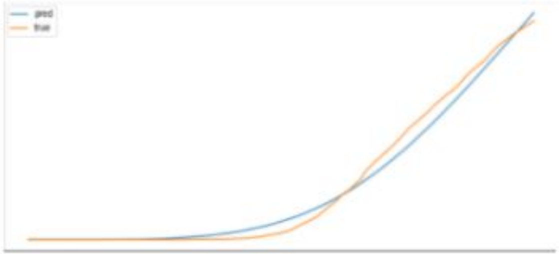
Bi-LSTM best MAPE

SEIR - HCD best MSLE

SEIR - HCD has good enough forecasting performance

Country	SEIR - HCD MAPE	LSTM MAPE	Bi-LSTM MAPE	SEIR - HCD MSLE	LSTM MSLE	Bi-LSTM MSLE
France	0.014	**0.002**	0.007	8.07	**5.24**	8.14
Spain	0.002	0.007	**0.0005**	**0.43**	8.36	2.13
Italy	0.06	0.001	**0.00026**	3.91	4.32	**0.46**
Germany	0.004	0.012	**0.0001**	**0.69**	5.8	1.09
United Kingdom	0.04	0.03	**0.006**	**1.783**	16.005	7.24
Means	0.026	0.011	**0.002**	**2.98**	7.95	3.819





Reproduction Rate Rt

Controls the Spread of the Virus

- How many individuals does an infected person transmit the virus to
- An Rt around 1 means the spread is under control
- An Rt highly over one indicates exponential spread growth

Measures taken by governments target the reduction of Rt

Stricter Measures assert a faster Rt reduction

Utility Function : Rt Reduction Rate

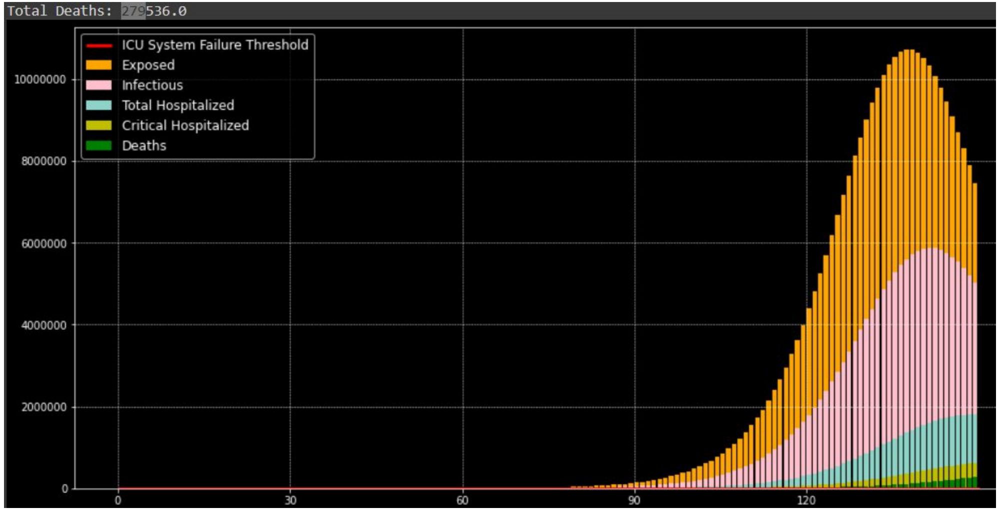
Assumption : Virus is solely transmitted from a person to person after they are in close contact

Rt Reduction Rate : New Rt = R0 - Reduction_Rate * R0

Reduction_Rate = (1-b) ^ (c*a)

- b: Transmission Rate -> Normal setting 0.45
- c: Average contracts per person per day -> Normal settings 13
- a: Government's strictness : Normal setting -> 0.9

Results without Intervention



In this nightmare scenario, without any interventions after 2 months the whole hospital system collapses and almost 300.000 people die in just a month

Intervention is necessary!!!

Intervention

User will be able to select the date and the duration of the intervention.

He will also be able to manipulate parameters a,b,c to choose the desired measures.

Different scenarios could be tested based on the parametric feature extraction.

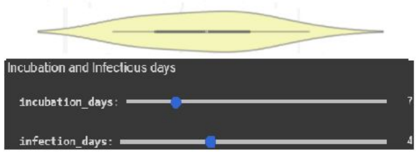
Intervention examples:

A. Total Lockdown (w. Obligatory Masks) : a = 0.5 , c = 3, b = 0.1
B. Work from home (w. Obligatory Masks outdoors) : a = 0.60 , c = 9 , b = 0.2
C. Closed Bars and work from home (w. recommendations for mask) : a = 0.7 , c = 5 , b = 0.2

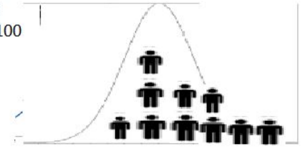
Each intervention works best under specific scenarios. These scenarios are controlled by R_0

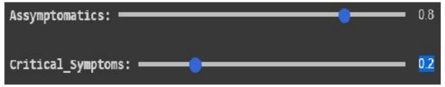
Framework Rough Draft

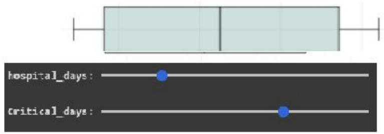
Pandemic Dynamic

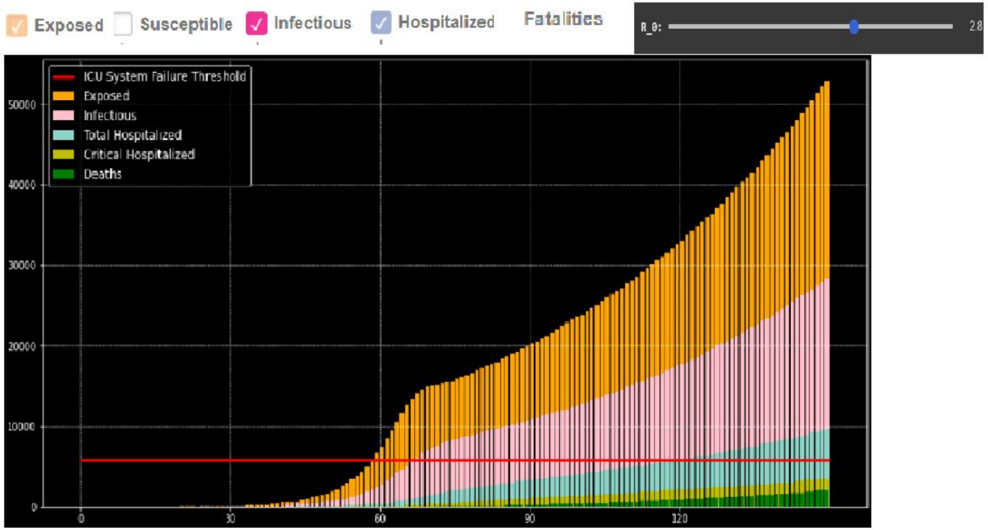


Probabilities:
Number per 100 individuals









Measures Taken:

Transmition Rate	Contracts Per Day	Guideliness Strictness
0.1	15	
Normal Setting : 0.45	Normal setting 15	Normal settings 0.9

Decision Making Process

Goal of the Framework:

Find the optimal measures combination for each scenario based on different parametric set ups.

Consider different scenarios

Multi-criteria Decision Making

- Long Term flat spread of the virus
- Avoid Hospital System Failing
- Don't abuse max strictness of measures to prevent socioeconomic consequences

Normative vs Descriptive Decision Making

DM process depends on policy makers.

- ➔ Smooth Spread of Virus without Hospital system Failing(Sweden)
- ➔ Avoid Loss(number of deaths) is prioritized over letting the virus smoothly spread in the community(Greece)

Recommendation:

Normative Decision Making

- Long Term reward:
  - Sweden Approach

Optimize Utility Function under each scenario to ensure long term gain