UCLA

CS145 Project Introduction

COVID19 Prediction

Instructor: Yizhou Sun

TAs: Junheng Hao, Shichang Zhang, Yue Wu, Zijie Huang

10/12/2020

UCLA

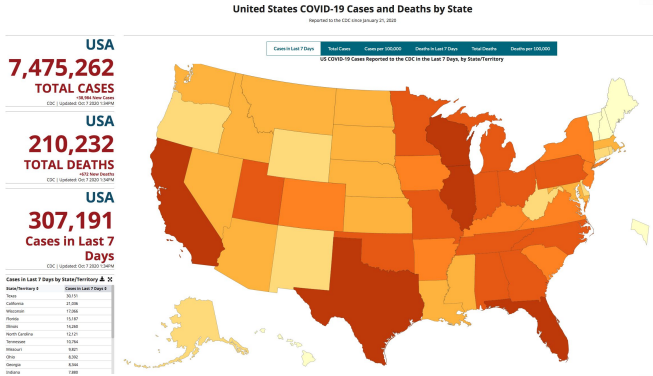
Project Introduction

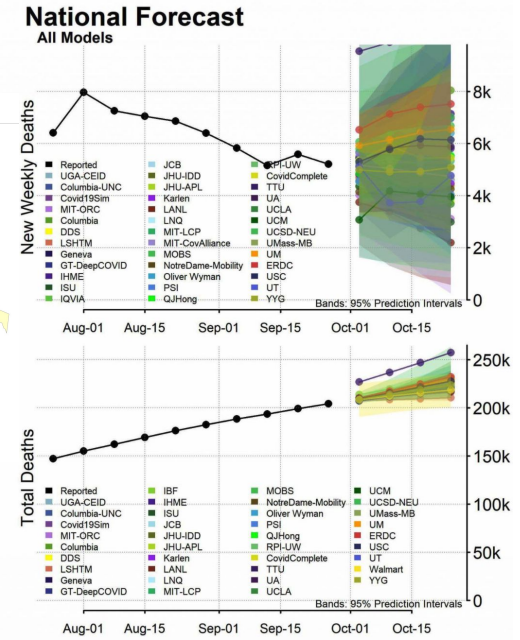
- Background & Motivation
- Project Task and Dataset
- Evaluation
- Project Deadlines and Grading

Background

COVID19 Prediction :

The rapid spread of COVID-19 has had and continues to have a significant impact on humanity. Accurately forecasting the progression of COVID-19 can help government monitor and take actions to combat it.





[1]https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

Background Motivation

- Based on various daily monitoring data of each **U.S. state** for a given time period (e.g. Apr-Aug), for an unseen time period (Sept), can you predict the daily **#case** and **#death** for each state?
- Timeseries Prediction with various types of data.
- A good fit for our class!

Province_State	Confirmed	Deaths	Recovered	Active
Alabama	3563	93		3470
Alaska	272	8	66	264
Arizona	3542	115		3427
Arkansas	1280	27	367	1253
California	22795	640		22155
Colorado	7307	289		7018
Connecticut	12035	554		11481

[1]https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data

Task

Based on the information from Apr.12 to Aug.31 of :

- Timeseries data for each state :
  - 10 features with full description on JHU_github
  - `Features:'Confirmed', 'Deaths', 'Recovered', 'Active', 'Incident_Rate', 'People_Tested', 'People_Hospitalized' ,'Mortality_Rate', 'Testing_Rate', 'Hospitalization_Rate'`
- Daily mobility data among different states [1]
- (Optional) Datasources can be added by yourselve :D (e.g. Placekey community data product)
  - Additional data can be used after permission by TAs. (Overall, any data that is befor Sep.01.2020 should be fine.）

[1]https://docs.safegraph.com/docs

UCLA

Task

Aim: Predict **#case, #death (cumulative value)** for each state from **Sep.1-26**:

- **Output 1:** Daily predicted # case, # death for each state
  - # of predication values: **26*50*2**
- **Output 2:** Daily predicted #case, # deaths on the final week data which would have ground truth only after you submitted your predictions. (can use data up to the prediction starting date to finetune your model.)
  - # of prediction values: **7*50*2**
- Ground Truths are accessible online for Output 1. DO NOT use them! (Test set leakage will be scored 0 for Output 1).
- We will test your model's performance on Output 2, also possibly reproduce you reported results for Output1 and Output2.

Task

$$APE = \frac{|Predicted - Truth|}{|Truth|}$$

How to evaluate:

- MAPE: mean absolute percentage error (take the average over all datapoints)
- Leaderboard ranking depends on Output1, but final projects score would depends on both Output 1 and Output 2.

Try your model on the Kaggle competition (limited 3 submissions per day):
https://www.kaggle.com/t/ff4c063c7b844ac29e5b709801766038

Submission file name: TeamNumber_Model.csv (e.g. Team1.csv)
More details read the information on Kaggle website.

UCLA

Project Grading (Total 25 Points)

- Midterm Report (**2** points)
- Final Report (**10** points)
  - Clairity in model explanation, different implemented model variants, etc.
- Performance on Kaggle (**13** points)
  - Evaluated by the results both from Output 1 and Output 2
  - Both MAPE score and rankings among all groups
  - Passing scores (~60%, 7 points) for models outperforming the given baselines; scores of most groups will range between 80%-100% (9-13 points).

UCLA

Project Group Formation

- Submit group information and register your group on Kagge by the end of Week 2.
- Team name, Group ID (will be assigned), member info (names, UIDs, emails)

UCLA

Project Midterm Report

- **Approximately 3 pages**
- Current progress about project, including
  - Data processing and transformation
  - Designed & tested models / methods
- Discussion and future project plan
  - Some conclusions and findings
  - Analysis of current models and techniques
  - Timeline of future project plan (around the next 4 weeks)

Project Final Report

- **No longer than 10-page PDF in ACM paper format:** https://www.acm.org/publications/proceedings-template
- Must include:
  - Group member information
  - Data selection and pre-processing
  - Model and techniques
  - Evaluation, observations and insights, conclusion
  - Current leaderboard rank and score
  - References and credit (papers, other's codes, maximum 1 page)
  - Related work (maximum ½ page)
  - Task distribution form
  - Peer evaluation form (separately submitted by individuals)
- Must NOT include:
  - Background or too much description on given original datasets
  - Any source code

UCLA

Task Distribution Form: Example

Task	People
Data processing	Student A
Implementation: Algorithm 1	Student B, C
Implementation: Algorithm 2	Student B, D
Implementation: Algorithm 3	Student A, D
Writing final report	Student C

Peer Evaluation Form: Example

CRITERIA	NAMES		
	John	Alice	Bob
Attendance at group meetings	4	4	3
Availability when needed	5	4	3
Highly contributed to writing and proof reading of the final report.	5	5	1
Reliability	5	5	2
Contributed ideas that were of high quality.	4	5	2
Approximately, the amount of time spent on this project was comparable to other group members.	5	5	2
Overall (Would you work with them again?)	5	5	2

Question:
Do you think some member in your group should be given a lower score than the group score? If yes, please list the name, and explain why.

Important Dates & Milestones

- **Oct.18**: Group formation due (link)
- **Nov. 9**: Midterm project report due
- **Dec.10**: Kaggle Submission Due (release new data for Output2 around a week before)
- **Dec.18**: Final project report due (together with all codes)

Note that the deadlines are subject to change according to the class schedule (avoid other deadlines of homework and exams).

UCLA

Q & A

UCLA

Thank you!

Enjoy "mining" and good luck!