An open toolbox for generating map of actively confirmed SARS-CoV-2 or COVID-19 cases in Vietnam

Duc Chung Tran
Computing Fundamental Department, FPT University, Hoa Lac Hi-Tech Park, Vietnam

ABSTRACT
The recent outbreak of novel coronavirus, SARS-CoV-2 or COVID-19, discovered in late 2019, being continued to spread across regions worldwide, has resulted in 1,914,916 “confirmed” cases with up to 123,010 deaths, as in situation report – 85 by World Health Organization (WHO). Most of the developed disease monitoring and tracking tools currently available only present the reported cases up to country-level and not detail down to provincial- or state-, city-level within the countries. This is insignificant for supporting activities in quickly reducing and preventing the spread of the disease within a certain country because further detail potential infectious locations are not provided for people to avoid traveling or passing by there. Thus, this work presents an open toolbox for generating map of actively “Confirmed” cases in a country, i.e., Vietnam, given a dataset containing their statuses and current locations, detail down to provincial or state-, city-level. The newly released algorithm reduced approximately 24.41% of processing time of the preceding one. In addition, the algorithm can be easily extended for supporting other countries given suitable datasets.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Duc Chung Tran,
Computing Fundamental Department, FPT University,
Hoa Lac Hi-Tech Park, Hanoi, 155300, Vietnam.
Email: chung.tranduc89@gmail.com

1. INTRODUCTION
The recent outbreak of coronavirus, SARS-CoV-2 or COVID-19 [1–7], discovered in late 2019, being continued to spread across regions worldwide [8-11], resulted in 1,914,916 “confirmed” cases with up to 123,010 deaths [12]. The disease is spreading extremely quickly resulting in exponentially infected cases all over the world [13-16]. As a result, catastrophic consequences happen to the world’s economics [13, 17-20]. Since the detection of the virus, researchers worldwide have paid rapid attentions and put great efforts in addressing plenty of socioeconomic consequences. In [15] presented an approach for predicting the trend of the pandemic in Italy with the main purpose for developing strategies for public health. The time series data from 22th Jan 2020 to 16th Mar 2020 were used by the model named eSIR. The model utilizes different intervention measures’ effects of dissimilar period in the prediction. In addition, Markov Chain Monte Carlo methods were used to estimate the basic reproductive number. Working on different approach, [14] discussed the severity of acute respiratory syndrome of infected patients and the challenges of this pandemic. Based on the research, the mean incubation period was found to be approximately 7 days while the basic reproduction number was between 2.24 and 3.58 [14]. Therefore, patients should be monitored closely. Aside from the pneumonia as the most manifestation of the virus, [4] suggested that extra-pulmonary symptoms such as initial cerebrovascular shall manifest.
the virus. In order to prepare for a quick response to COVID-19, UW Medicine presented an impatient response plan for high-quality palliative care [8] while Australian authority kept continuous updates of the situation in its formal reports [3]. In another research, [13] discussed the implication for physiotherapy topic with respect to global impact. It was stated that the clinical health system has plenty of difficulties to deal with the surge; and at global level, workers are among the ones to have least access to such medical care system. Meanwhile, it was studied that pregnant shall expose to greater risks than normal people when infected by the virus [2]. For supporting fast detection of the virus, i.e., under 30 minutes, [9] presented an approach by reversing transcription-loop-mediated (RT-LAMP) isothermal amplification.

It is seen from the above analysis that various research topics were studied since the pandemic of COVID-19. In order to effectively prevent the spread of the disease, it is critically important for people in a specific country to be aware of the potential infectious locations (i.e., provinces or states or cities) to avoid traveling or passing by there. There have been multiple attempts to present the global map of the infectious condition. However, currently there is only country-level data in reports of WHO [12], real-time dashboard COVID-19 tracking [1] as shown in Figure 1, dashboard in [1, 21], information in [22-31] as shown in Figures 2, 3. Meanwhile, data in [32] had provincial level for Vietnam, however they are incomplete, only for some cities such as Hanoi, Vinh Phuc, Nha Trang and Ho Chi Minh. Thus, the provided information is not sufficient for offering essential warnings for people living in the country to be aware of the potential infectious areas and avoid to move there for preventing the spread of COVID-19. Here, it should be noted that the potential infectious areas are defined as those the actively “confirmed” cases have passed by recently which is detailed down to district or state level.

Figures 1, 2 and 3 show different forms of presentation of COVID-19 data, but none of them clearly reveals the potential infectious areas at the state or city level. It is seen that, although the infection spread has extended to many provinces and cities in Vietnam, the data are not sufficient to provide warnings for people living in the country. Therefore, this work addresses the current issue of the country-level COVID-19 information in the existing disease monitoring platforms, particularly for the case of Vietnam. Below are the contributions of this research paper:

a. It does take into consideration of deeper information level, to be particular, state- or city-level in developing the “confirmed” cases map.

Based on the analysis, it is found that the obtained country-level information [1, 22-25] is only useful for statistics analysis purpose. Whereas in deeper look, state- or city-level are much more useful and meaningful for the living people in the country. The key reason is that without knowing information from a city with newly “confirmed” cases, many people may pass by the city or the infectious area which can spread the virus. Thus, for the safety of the country, it is important to keep track of all “confirmed” cases and provide timely information to warn citizen for self-avoiding or preventing the spread of this deadly virus. Therefore, this work addresses the current issue of the country-level COVID-19 information in the existing disease monitoring platforms, particularly for the case of Vietnam. Below are the contributions of this research paper:

a. It does take into consideration of deeper information level, to be particular, state- or city-level in developing the “confirmed” cases map.

An open toolbox for generating map of actively confirmed SARS-CoV-2 or COVID-19... (Duc Chung Tran)
b. The demonstration research for Vietnam; it can be easily extended to other countries since the complete source code, demonstrated kernel [33] and sample dataset [34] were released publicly.

c. The solution (the first Open Toolbox for Generating SARS-CoV-2 or COVID-19 Map of Actively Confirmed Cases in Vietnam) is completely free since it does not utilize any commercial cloud application programming interface (API) services from common providers such as Google, Microsoft. Thus, this enables wide use of the work even in undeveloped countries with low incomes where access to such premium information is limited.

![Figure 2. 94 Confirmed COVID-19 cases in Vietnam, without detail locations, at 2 AM, 22 March 2020, GMT +7 [22]](image)

![Figure 3. 94 Confirmed COVID-19 cases in Vietnam, without detail locations, at 2 AM, 22 March 2020, GMT +7 [24]](image)

2. RESEARCH METHOD

2.1. Data preparation

In this work, in order to generate a map of the actively “Confirmed” cases in Vietnam, the input data were manually processed from reliable sources of information including release news of Ministry of Health (MOH), Vietnam [35], local news websites such as VnExpress.net [36], DanTri.com.vn [37], ThanhNien.vn [38], and TuoiTre.vn [39]. Here, it should be noted that the news from local websites were also sourced and summarized from MOH thus, they had equivalent information publishing credential as the MOH.

Although the dataset contained various information fields, in order to generate the actively “Confirmed” cases map, the information from column “Case” (which contains patient identification, anonymous abbreviation), “Current Location” (which contains the currently district/city/state that the patients are located), “Confirmed” and “Recovered” statuses were used. When a case is “Confirmed”, his status will be marked as 1, else it will be marked as 0. When a case is “Recovered”, his status will be marked as 1, else
An open toolbox for generating map of actively confirmed SARS-CoV-2 or COVID-19... (Duc Chung Tran)
to the state name, this character is stripped from the variable `state` forming the needed information to fill up “State” column.

```python
for i in range(nrow):
    addr = dfState.iloc[i,0]
    if (str(addr) == 'nan'):
        print('index = ', i, ' addr = ', addr, ' -> no address')
    else:
        s = addr.split(',')  # delimiter = ','
        state = s[len(s)-2]  # get province / state
        state = state.strip()
        dfState.at[i,'State'] = state
        print('state = ', dfState.iloc[i,1])
```

Figure 6. Algorithm to retrieve “state” from “current location”

Finally, the algorithm loops through all rows in dataframe “dfi” and searches for the current latitude, longitude of each patient, then adds them to the plotting map. Because there shall be a location having more than one “Confirmed” cases, multiple searches of the same location is not necessary. Thus, to implement one-time search only for each location, algorithm in Figure 7 is used.

```python
for i in range(nrow):
    addr = dfi.iloc[i,1]
    if (str(addr) == 'nan'):
        print('no address')
    else:
        # search if address appear once, lat long are available already, then no need to retrieve lat long again
        alreadyExist = 0
        foundPos = 0
        for j in range(0, i-1):
            addr2 = dfi.iloc[j,1]
            if (str(addr) == str(addr2)):
                alreadyExist = 1
                foundPos = j
                break
        if (alreadyExist):
            lat = dfi.iloc[foundPos,5]  # col Lat
            long = dfi.iloc[foundPos,6]  # col Long
        else:
            try:
                lat, long = latlongGet(addr)
            except:
                print('')
            if(lat):
                dfi.iloc[i,5] = lat  # col Lat
            if(long):
                dfi.iloc[i,6] = long  # col Long
            # print(lat, long)
```

Figure 7. Reducing multiple redundant latitude, longitude searches of “current location”

For each location in dataset at row $i^{th}$, a time sleep of 1 second is added to ensure stability of information transferred between client and the licence-free map server. If there is no address information, the algorithm prints a log indicating there is no address in the input field, else, it starts the search for latitude and longitude. Two intermediate variables called `alreadyExist` and `foundPos` are created. The former is used to indicate that the location details already exist while the latter is used to locate the position (row number) of the details stored in dataframe `dfi`. An internal dataframe search is performed to find if the address appears in the preceding rows $j^{th}$ ranging from 0 to $i-1$. If the address is found, the algorithm sets `alreadyExist = 1` and `foundPos = j`, then breaks the loop, directly retrieves latitude and longitude
information to fill up the $i^{th}$ rows in the respective columns. If the address is not found, the latitude and longitude are retrieved by sending a query to the map server. By having this approach, multiple redundant queries between client and map server is greatly reduced, thus improving significantly processing time: reducing from 156.5 seconds [44] to 118.3 seconds [33] (approximately 24.41% reduction). Here it should be noted that, in order to position the country, i.e., Vietnam, into the center of the map, coordinates of Hue city, Vietnam is used as the map center. The minimum working example of the algorithm can be found in [33].

3. RESULTS AND DISCUSSION

Figure 8 presents the improvement of approximately 24.41% in processing time when plotting actively “Confirmed” cases map by reducing redundant latitude, longitude queries to map server when running the algorithm shown in Figure 7.

![Figure 8](image)

In addition, Figure 9 illustrates an output of the developed toolbox which reveals an insight into the potentially infectious areas within Vietnam. The time taken to run the developed toolbox is 118.3 seconds (version 8 of 8). In this figure, one easily observes that the COVID-19 spread to major center parts of North, Middle, and South Vietnam. The states that were in between had no indication of “Confirmed” cases. Thus, people should be more cautious when travelling in locations flagged in “Red”. When mousing over the highlighted location, information of actively “Confirmed” cases and “Recovered” cases are displayed in detail.

![Figure 9](image)
4. CONCLUSION

In conclusion, this work has presented a generic toolbox for creating a map of actively “Confirmed” cases in a country, i.e., Vietnam. It is evident that the algorithm’s processing time improved significantly (approximately 24.41%) by reducing redundant latitude and longitude queries sent to the map server. Although the tested dataset is for Vietnam, the algorithm can utilize similar datasets from any other countries for plotting the COVID-19 maps across those countries. Thus, it is definitely possible for the algorithm to generate all locations of the actively “Confirmed” cases in timely manner for any country given a suitable dataset. Future work will take into consideration of applying the toolbox to other nearby countries.

ACKNOWLEDGEMENTS

The authors would thank FPT University for supporting this research.

REFERENCES


An open toolbox for generating map of actively confirmed SARS-CoV-2 or COVID-19... (Duc Chung Tran)