CANADA COVID-19 TREND ANALYSIS

Manpreet Nanreh, Abdolreza Abhari
Department of Computer Science
Ryerson University
350 Victoria Street, Toronto, ON, M5B 2K3
mnanreh@ryerson.ca, aabhari@ryerson.ca

ABSTRACT

In this paper, the COVID-19 dataset from the Johns Hopkins CSSE (2020) was analyzed with the goal to perform trend analysis and determine whether the data follows an exponential trend or a logistic trend. Through curve fitting, it was determined that logistic curve fits the growth more than the exponential curve.

Keywords: COVID-19, data analytics, curve fitting

1 INTRODUCTION

COVID-19 has had a great impact all over the planet and it is spreading at a rapid rate in most countries. As of April 23, 2020, there have been total of 43,286 confirmed cases of COVID-19, 14,761 recovered cases of COVID-19 and 2,240 fatal cases of COVID-19 in Canada as reported by Johns Hopkins CSSE (2020). On March 11, 2020, the World Health Organization declared COVID-19 as a worldwide pandemic as indicated by the Infection Prevention and Control Canada (2020). This paper focuses on discussing the trend analysis this pandemic follows for confirmed cases in Canada.

2 RELATED WORK

In the work done by Rozema (2007), the technique of curve fitting was used in order to find the best fit logistic curve for the trend analysis of SARS and Measles outbreaks. This inspired to apply the similar type of analysis for COVID-19 and determine whether the pandemic is following an exponential or a logistic trend. From that paper at page 250, the equation for logistic curve used for this study was as follows:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \qquad \textit{Eq. 1}$$

The parameters in *Eq. 1* are explained as follows: L refers to the population to be infected, k refers to the initial growth rate and x_0 is the inflection point for the logistic curve.

3 METHOD AND DATASET

The dataset used in this study was obtained from the Johns Hopkins CSSE (2020) which is updated on daily basis and is provided in time series format. For this study, the data for all the Canadian provinces was combined such that analysis could be done for the entire country. The data was interpreted from the first day the cases were reported which in this dataset occurred on January 20, 2020 and that was marked as Day 0. The exponential and logistic curves were fit to the actual data for confirmed cases in Canada. The following equation for exponential curve was used

$$f(x) = ae^{bx+c} \qquad \textit{Eq. 2}$$

The equation for logistic curve is listed in *Eq. 1*. These were fit to the data using the curve fit function from the SciPy library written in Python by Virtanen et al (2020). The library optimizes the fit by minimizing the least squares error. The library required an initial guess to be supplied for each of the parameters in order to achieve convergence and determine the optimal parameter values. For the logistic curve the initial guess for parameters was picked as follows: L was picked as the maximum value of the data, k started at the value 1 and x_0 was picked as the mean of the data. For exponential curve, the initial guess for parameters a and b was picked as the value 1 and the parameter c was picked to be 12. After the SciPy package optimized the curve fit, it provided the best fit parameter values. The measurement metric used to compare the curves was mean squared error and the R^2 value. For the logistic curve, the inflection point was determined by computing the second derivative of *Eq. 1* and evaluating it when it equals zero. As Rozema (2007) states in the paper that the inflection point for *Eq. 1* is found when $x = x_0$.

4 RESULT AND DISCUSSION

The curve fit for *Eq. 1* and *Eq. 2* can be visualized in **Figure 1**.

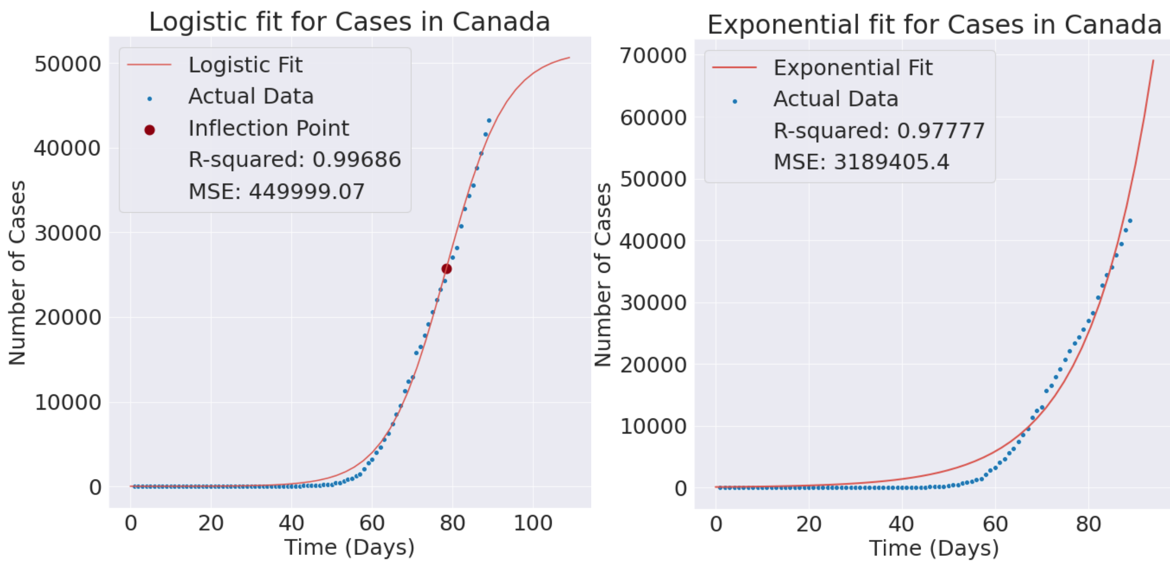

Figure 1 The graph on the left shows the logistic fit with the inflection point for the confirmed cases and the graph on the right shows the exponential fit for the confirmed cases in Canada.

From the analysis, it was found that the logistic curve fit had a lower mean square error and a higher R^2 than exponential curve, therefore, it is a better fit. Their equations with the parameters were found as:

$$f(x) = \frac{51474.1308}{1 + e^{-0.134678860(x-78.4386535)}} \qquad \textit{Eq. 3}$$

$$f(x) = 4.92523249e^{0.0728305x+2.7029153} \qquad \textit{Eq. 4}$$

The inflection point for the logistic curve was found to occur at day 78 which refers to April 13, 2020 until which 25,667 confirmed cases had been reported thus far. This means that until April 13, 2020 the trend the number of confirmed cases were following was that of an exponential nature, but after that date the trend changed to logistic nature.

5 CONCLUSION

This paper discussed the type of trend the COVID-19 pandemic is likely to be following for confirmed cases as of April 23, 2020. Through the curve fit of exponential and logistic functions, it was found that the growth of confirmed cases is most likely to be following logistic trend and indicates that Canada may have passed the inflection point and is on the path to flattening the curve.

REFERENCES

Edward Rozema 2007. *Epidemic Models for SARS and Measles*. The College Mathematics Journal, 38:4, 246-259, DOI: 10.1080/07468342.2007.11922245

Infection Prevention and Control Canada. 2020. *Coronavirus (COVID-19)*. Infection Prevention and Control Canada. https://ipac-canada.org/coronavirus-resources.php.

Johns Hopkins CSSE. 2020. *2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository*. Johns Hopkins CSSE. https://github.com/CSSEGISandData/COVID-19.

Pauli Virtanen, Ralf Gommers, et al. 2020. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods 17 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2.