Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(4): 4572–4585.
DOI: 10.3934/mbe.2021232
Received: 21 March 2021
Accepted: 17 May 2021
Published: 26 May 2021

Research article

Track the dynamical features for mutant variants of COVID-19 in the UK

Ray-Ming Chen*

Department of Mathematics and Statistics, Baise University, 21 Zhongshan No.2 Road, Baise 533000, China

* **Correspondence:** Email: raymingchen@bsuc.cn.

Abstract: *Aims*: The main purpose of this study is to explore whether the new variant SARS-CoV-2 VOC 202012/01 in the UK is equipped with some leading or underlying features. *Methods*: We apply a systematic and persuasive approach to reveal the underlying dynamical features of this variant. The approach utilises extracting the main features, which consist of 3-valued features, via the time-series data for new cases, 28-day deaths and 60-day deaths. The experimental samples chosen rely on the the rolling sets of regional data vectors whose dimensions are all 7 days. These data sets are projected onto the 3-valued features to yield the vector rejections. Then the minimal features are thus extracted by the minimal total norms. Then we map out the traces of the similarities between all the extracted time-varying features. *Results*: Our findings, no matter in preliminary or follow-up study, clearly show there is no consistent and substantial shift in the 3-valued features even after the occurrence of this new variant - this might validate the efficacy of the current vaccines against this variant. *Conclusions*: Since the underlying features of the mutant is unchanged, and the leading feature of B1.1.7 is not yet present, it might help us make the lockdown decision "choose the lesser of the two evils: pandemic and economic woes", and validate vaccine developments or adopt preventive measures.

Keywords: mutant variants; UK COVID-19; distance tensor product; 3-valued features; feature extraction

1. Introduction

In 14th December 2020, the UK has found a new variant of SARS-CoV-2 or SARS-CoV-2 VOC 202012/01 (Variant of Concern, year 2020, month 12, variant 01) [1–3] or B.1.1.7. There are some researches on the emergence of this SARS-CoV-2 variant [4, 5]. The variant [6] has caused a huge concern across the globe [7] and has spread across other continents. Based on known epidemiologic, scientific modellings, and clinical findings, it suggests this variant has indeed increased the overall transmissibility.

On the other hand, it also indicates that there is no change in severity [8] - measured by length of hospitalization and 28-day case fatality - or occurrence of reinfection in the UK [9], despite some scientists' dire warnings [10]. Even though the evaluation of variant is still going on [11], we need to find a systematic method to delve into the fundamental features of this mutant variant. This might concern whether 202012/01 diminishes the potency of the developed vaccines [12].

To trace the fundamental features of this new variant, we sample 9 regional daily COVID-19 data in the UK from 27th March, 2020 to 8th, January, 2021 [13] as the preliminary study and sample data from 28th July 2020 to 11th May 2021 as the follow-up study. Then we create a set of batches of data vectors whose dimensions are decided by the days chosen. In our research, we choose 7 days as the datum dimension. Furthermore, to find the fundamental features of B1.1.7, we introduce the 3-valued featured vectors whose elements consist of only -1, 0 and 1. These features are easy to apply and would serve as our fundamental database for further feature recognition. We then utilise the concept of vector rejection to measure the similarities between the set of data vectors and the features based on the total norms for the vector rejections. Afterwards, we obtain distance tensor products which record such similarities. By choosing the minimal total norms regarding the features, we retrieve the minimal total norms and their corresponding features. Since these extracted features vary, we trace their paths via cosine values for the dynamical features. This path would reveal the fundamental trend and dynamical structures of the features for B1.1.7 in the UK. The preliminary focuses on comparing the dynamical structures of COVID-19 with respect to data with London (9 regions), where it is supposed to be the onset of B1.1.7., and without London (8 regions).

Our results show there is no obvious shift for the main features before and after the variant - this confirms there is no fundamental change in the behaviour of the variant. These findings indicate B1.1.7 contain no leading features, and, in principle, shall bolster the efficacy of the vaccines.

2. Methodology

We devise rejection distance which derives from the concept of vector rejections and three-valued featured database. Both shall serve as our fundamental methods in tracking the features of B1.1.7.

2.1. Distance tensor product

Claim 1. *(rejection distance) The norm of $\vec{c}$, a vector rejection of $\vec{a}$ on $\vec{b}$, is* $||\vec{c}|| = \frac{\sqrt{(||\vec{a}|| \cdot ||\vec{b}||)^2 - (\vec{a} \cdot \vec{b})^2}}{||\vec{b}||}$ *or specifically the rejection distance between $\vec{a}$ and $\vec{b}$*

$$RejDist(\vec{a}, \vec{b}) := \frac{\sqrt{(||\vec{a}|| \cdot ||\vec{b}||)^2 - (\vec{a} \cdot \vec{b})^2}}{||\vec{b}||}.$$

Proof. By the property that $\vec{c} = \vec{a} - ||\vec{a}|| \cdot cos\theta \cdot \frac{\vec{b}}{||\vec{b}||} = \vec{a} - \frac{\vec{a} \cdot \vec{b}}{||\vec{b}||^2} \cdot \vec{b}$ and the operation of the inner product $\cdot$, the result follows immediately. □

Rejection distance acts as a role for measuring the distances between features, including empirical features or theoretical features (3-valued features, for example). This measurement will serve our optimal feature findings.

Definition 2.1. *(Distance Tensor Product, DTP) For any ordered set of vectors, whose dimension are the same,* $V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m)$ *and* $W = (\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_n)$*, we define their distance tensor product by a m-by-n matrix* $DTP(V, W)$ *whose* (i, j) *element is* $RejDist(\vec{v}_i, \vec{w}_j)$*.*

DTP shows the norms of the vector rejections, which reveal the distances between a set of given datum vectors and the featured vectors or the distance between feature vectors.

2.2. 3-valued features

Unlike Principal Component Analysis, or PCA [14], which heavily relies on the statistical property, the 3-valued features are applied directly to trace the trend of mutant variants of COVID-19. 3-valued features consist of a set of value $\{-1, 0, 1\}$, which is also used in wavelet analysis [15]. In total, there are $3^n - 1$ features (or vectors) - the 0 vector is excluded from the whole potential features, where n denotes the dimension of the datum vector.

3. Implementation procedures

3.1. Preliminary settings

In this section, we specify the implementation processes mentioned in section 3. From the database, we select three metrics: newCasesBySpecimenDate (M1, or 1), newDeaths28DaysByDeathDate (M2, or 2), and newDeaths60DaysByDeathDate (M3, or 3) - there indexes 1,2,3 will be used later on; and Area Type "Region" (there are 9 regions: Yorkshire and The Humber (R1), East of England (R2), North East(R3), North West(R4), South East(R5), East Midlands(R6), South West(R7), London(R8), West Midlands(R9)). The data are sorted by date, from 27th, March, 2020 to 8th, January, 2021. The data are stored in three matrices, according to M1, M2, and M3. The we calculate their first difference (change) of the new cases, 28-day deaths, and 60-day deaths, as shown in the following (the length for each vector is 287, since we sample 288 points of time). The value in interval i is the original value in $i + 1$ minus the one in i. When compared with the original data, these values are the second-order difference.

3.2. Data description

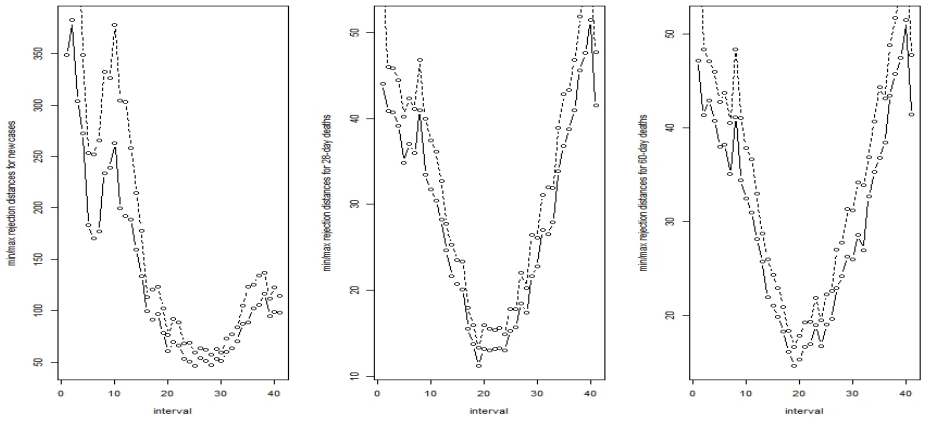
The data come from two main parts. One is for the preliminary study for the onset of the variant spread in the UK. The periods cover from 27th, March, 2020 to 8th, January, 2021. It lists the daily new cases, 28-day deaths and 60-day deaths of COVID-19 for 9 regions in the UK. Specifically, the metrics (or indicators) from the database are M1, M2, and M3. M2 is a variable for the new increase of people who had a positive test result for COVID-19 and died within 28 days of the first positive test, up to the date of death. M3 is a variable for the new increase of people who had a positive test result for COVID-19 and died within 60 days of the first positive test, up to the date of death.

This imported data are further separated into two sets of data - the preliminary study - one covers only 8 regions by excluding London and the other covers 9 regions which contain London. The second part is a follow-up study covering a period from 28th, July, 2020 to 11th, May, 2021 [16]. The data are not separated and are treated as a whole, i.e., the 9 regions.

3.3. Procedures

Then we set every 7 days as the dimension of one datum vector and 9 regions as the number of data vectors. In other words, there are $287/7 = 41$ ordered batches of data sets - each set contains 9 points (or datum sub-vectors, or simply datum vectors), and each point is a 7-dimensional datum vector. We form these data with respect to cases, 28-day deaths and 60-day deaths for the 287 intervals. There are several steps in the implementation. These procedures are designed for preliminary study, but the follow-up study would also follow suit.

1. Download the data related to daily COVID-19 cases and deaths.
2. Sort the data by the 288 dates, 9 regions and 3 metrics, according to our analytical purposes. The results are presented in Table 1. Let us still call them raw data, though they are actually processed to fit our purpose.
3. Compute the 287 first-order difference of the 288 time-series data, according to the 9 regions and 3 metrics. This will reveal the accelerated spreading trend of B.1.1.7. Let us call them difference data. Then the data are further extracted by regions and we name them difference datum vectors. The results are presented in Appendix B: difference datum vectors.
4. Split each 287 difference datum vector into 41 sub-vectors - each of which contains 7 elements (or daily data). Let us use $\{\vec{R}_{ij}^k\}$ to denote the j'th 7-day sub-vector for region i, where $k \in \{1, 2, 3\}$ ($k = 1$ to denote "cases", $k = 2$ to denote "28-day deaths", and $k = 3$ to denote "60-day deaths"), $1 \le i \le 9$ and $1 \le j \le 41$. For example $\vec{R}_{11}^1 = (54, 21, 57, 34, -22, -55, 51)$.
5. Specify the 3-valued features (7-element vectors), which is described in section 2.2, as our potential feature database. For 7-day period, there are $3^7 - 1 = 2186$ three-valued features. The one being excluded is the zero vector. The resulting feature database goes as follows ($\vec{f}_i$ stands for i'th feature):
   $\vec{f}_1 = (-1, 1, 1, -1, -1, -1, -1)$; $\vec{f}_2 = (0, 1, -1, -1, -1, -1, -1)$;
   $\vec{f}_3 = (1, -1, -1, -1, -1, -1, -1)$; $\vec{f}_4 = (-1, 0, -1, -1, -1, -1, -1)$; $\cdots$; $\vec{f}_{2183} = (1, 0, 1, 1, 1, 1, 1)$;
   $\vec{f}_{2184} = (-1, 1, 1, 1, 1, 1, 1)$; $\vec{f}_{2185} = (0, 1, 1, 1, 1, 1, 1)$; $\vec{f}_{2186} = (1, 1, 1, 1, 1, 1, 1)$. Let us use FDB to denote the 3-valued feature database, i.e., $FDB = \{\vec{f}_1, \vec{f}_2, \cdots, \vec{f}_{2186}\}$.
6. Calculate the distance tensor product (DTP) which is defined in Definition 2.1, where each element is defined by
$$DTP_{i,j,h}^k := RejDist(\vec{R}_{ij}^k, \vec{f}_h),$$
   where $1 \le i \le 9; 1 \le j \le 41; 1 \le h \le 2186$, and $k \in \{1, 2, 3\}$. The generated DTP could be regarded as three (regarding k) three-dimensional matrices (regarding i, j, h). The results are presented in Table 2.
7. Compute $DTP_{j,h}^k = \sum_{i=1}^{9} DTP_{i,j,h}^k$, for all $1 \le j \le 41$ and $1 \le h \le 2186$. The results are shown in Table 3.
8. Compute $max_j^k = max\{DTP_{j,h}^k : 1 \le h \le 2186\}, min_j^k = min\{DTP_{j,h}^k : 1 \le h \le 2186\}, argmax_j^k := argmax\{DTP_{j,h}^k : 1 \le h \le 2186\}, argmin_j^k := argmin\{DTP_{j,h}^k : 1 \le h \le 2186\}$ for all $1 \le j \le 41$. The results are shown in Table 4. Meanwhile, plot $\{min_j^k : 1 \le j \le 41\}$ and $\{max_j^k : 1 \le j \le 41\}$ for $k \in \{1, 2, 3\}$ as presented in Figure 1.
9. Compute $\{cos(argmin_v^k, argmin_{v+1}^k) : 1 \le v \le 40\}$ and $\{cos(argmax_v^k, argmax_{v+1}^k) : 1 \le v \le 40\}$ for $k \in \{1, 2, 3\}$, i.e., the similarity indexes (cosine values) between the extracted minimal



;

Figure 1. Min/max rejection distances between datum vectors and featured vectors for new cases (left), 28-day deaths (middle) and 60-day deaths (right).

(or maximal) features to yield a dynamical trend of the representative features. The results are presented in Figures 2–4.

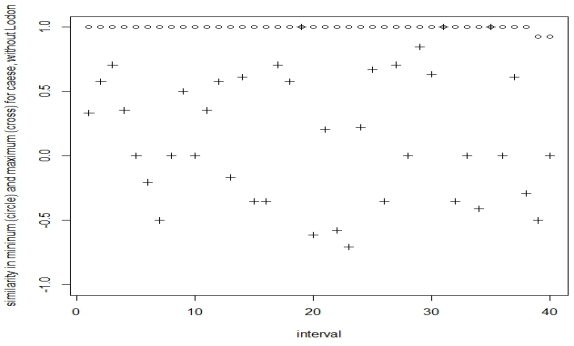
4. Results

In this article, we explore the trend of underlying features of the COVID-19 new cases, 28-day deaths and 60-day deaths to reveal whether the new mutant variants of COVID-19 in the UK have caused fundamental features, which are based on 3-valued featured vectors. The results are presented in two parts: preliminary study and follow-up study.

4.1. Preliminary results

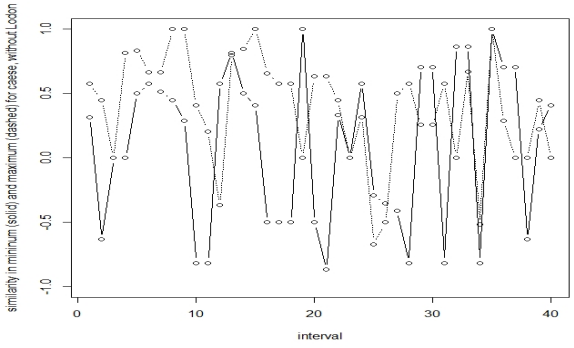
It covers the period from 27th, March, 2020 to 8th, January, 2021. The onset of the mutant variant spreads mainly in London. By comparing the dynamical trend of time-series of COVID-19 with London and without London, we shall see the role of B1.1.7 in the preliminary stage.

1. From Figure 1, we observe that the maximal and minimal rejection distance go almost together – an indication that the features chosen are qualified to capture their B.1.1.7's dynamical behaviours. Furthermore, no matter in the new cases, 28-day death, or 60-day deaths, when the time mutant variant takes place (around the tail of the three graphs), the diluting features are reverted in comparison to previous intervals. This indicates the new mutant variant changes the usual course of the collective viruses. In some sense, at this stage, it has not yet created a lead role in the pandemic, but it did affect the course of the development.
2. From the left-hand sides of Figures 2–4, the raw data for the minimal similarities almost reach 1, while the maximal similarities are pretty random. This indicates the optimal features are representative of the behaviour of B.1.1.7.
3. From the right-hand side of graphs in Figures 2 and 3, there is no obvious shift in the features when comparing the eight regions (without London) and the 9 regions (with London). This preliminary indication shows there is no clear feature from B.1.1.7 in terms of leading the pandemic. Since



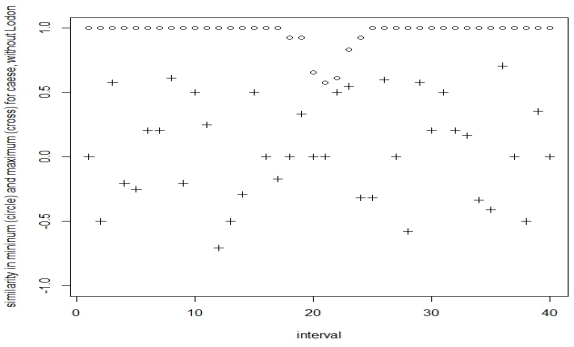
;

(a) raw data for minimal (circle) and maximal (cross) similarities for cases, excluding London



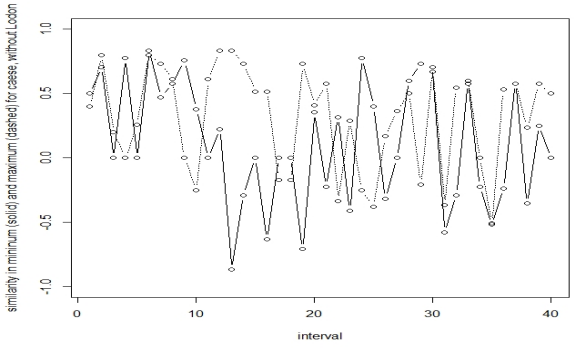
;

(b) difference data for minimal (solid) and maximal (dashed) similarities for cases, excluding London



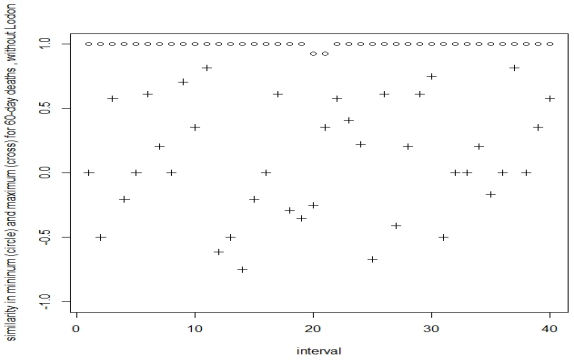
;

(c) raw data for minimal (circle) and maximal (cross) similarities for 28-day deaths, excluding London



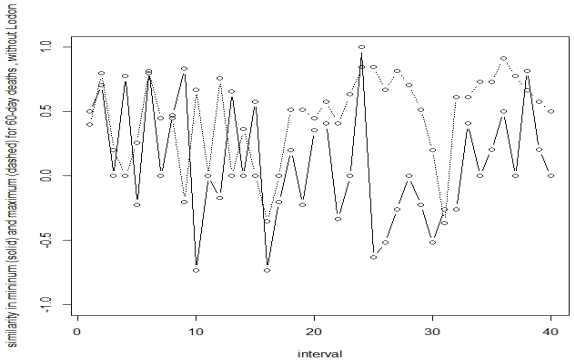
;

(d) difference data for minimal (solid) and maximal (dashed) similarities for 28-day deaths, excluding London



;

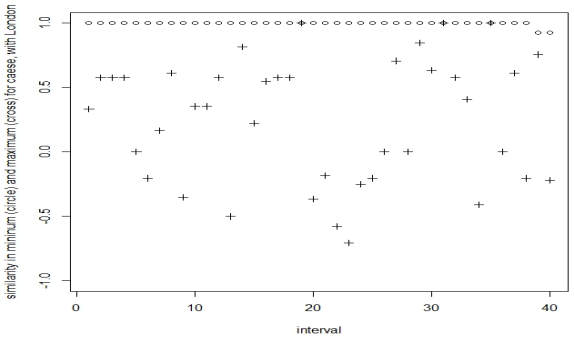
(e) raw data for minimal (circle) and maximal (cross) similarities for 60-day deaths, excluding London



;

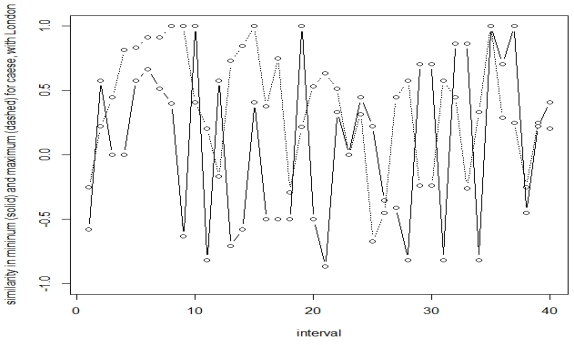
(f) difference data for minimal (solid) and maximal (dashed) similarities for 60-day deaths, excluding London

Figure 2. Minimal and maximal raw and difference trends without London.



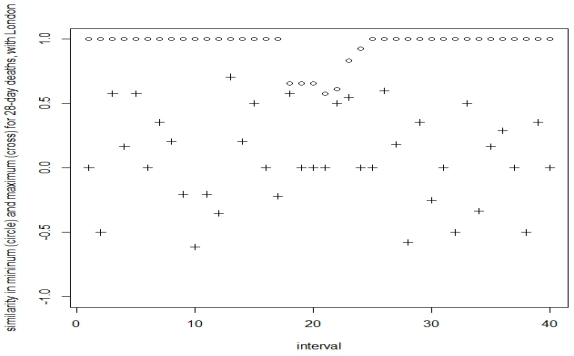
;

(a) raw data for minimal (circle) and maximal (cross) similarities for cases, including London



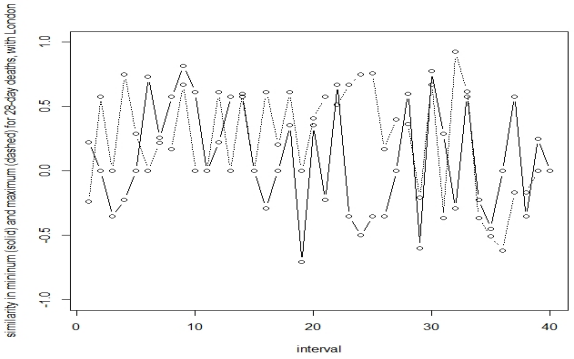
;

(b) difference data for minimal (solid) and maximal (dashed) similarities for cases, including London



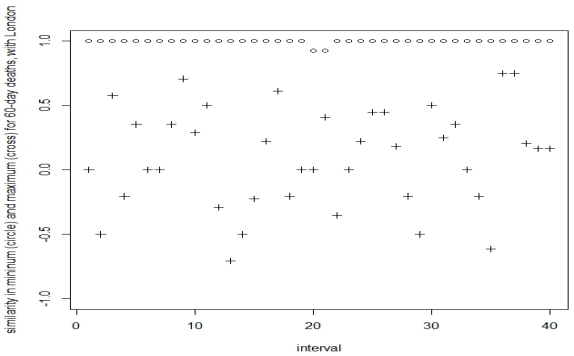
;

(c) raw data for minimal (circle) and maximal (cross) similarities for 28-day deaths, including London



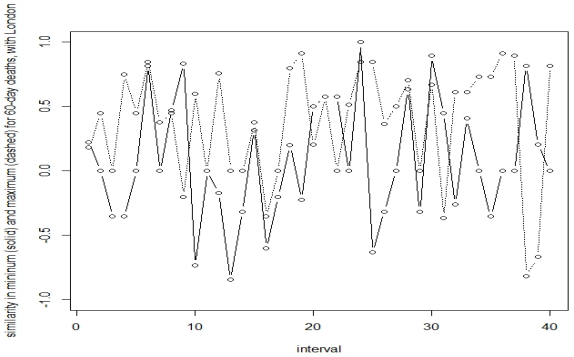
;

(d) difference data for minimal (solid) and maximal (dashed) similarities for 28-day deaths, including London



;

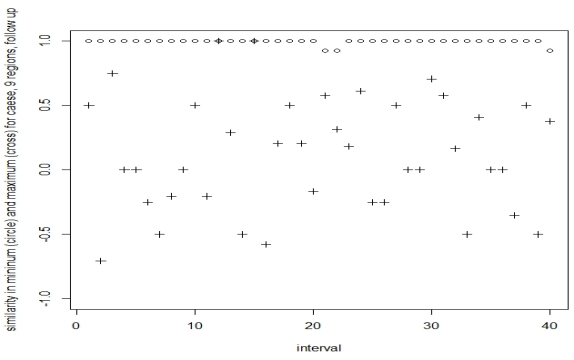
(e) raw data for minimal (circle) and maximal (cross) similarities for 60-day deaths, including London



;

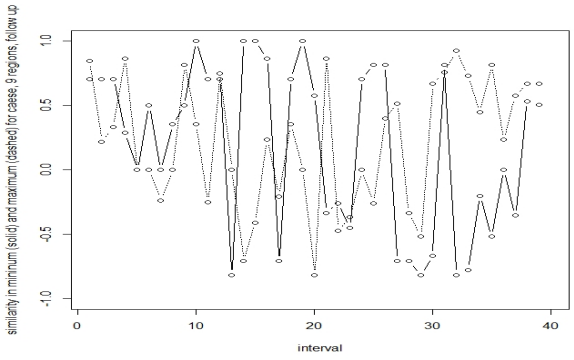
(f) difference data for minimal (solid) and maximal (dashed) similarities for 60-day deaths, including London

Figure 3. Minimal and maximal raw and difference trends with London.



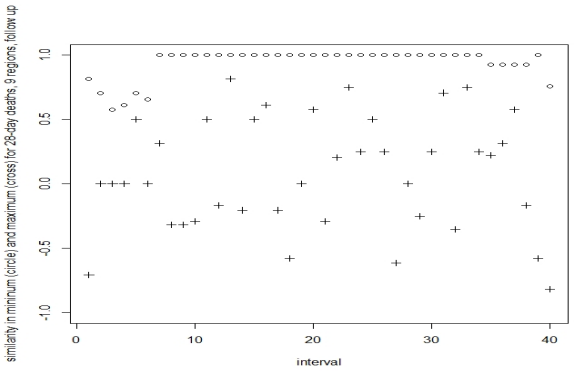
;

(a) raw data for minimal (circle) and maximal (cross) similarities for cases, 9 regions, follow up



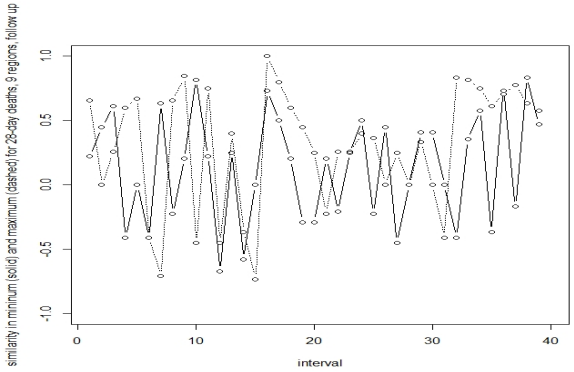
;

(b) difference data for minimal (solid) and maximal (dashed) similarities for cases, 9 regions, follow up



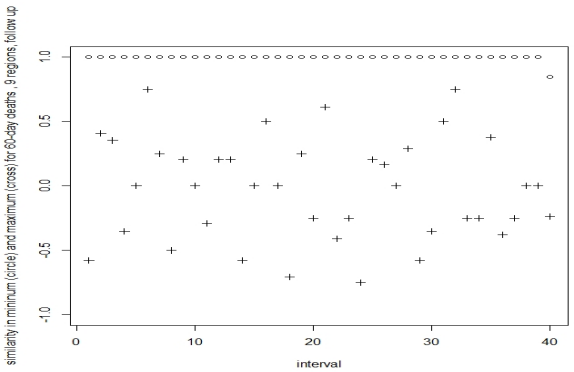
;

(c) raw data for minimal (circle) and maximal (cross) similarities for 28-day deaths, 9 regions, follow up



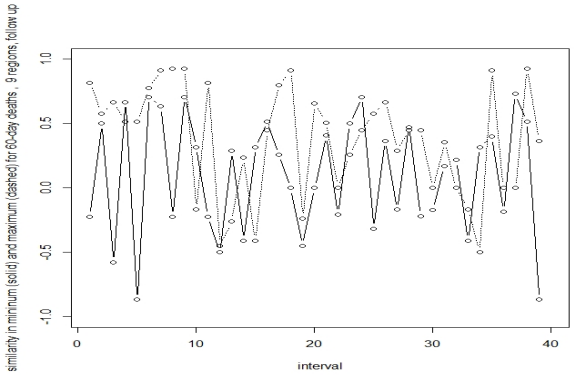
;

(d) difference data for minimal (solid) and maximal (dashed) similarities for 28-day deaths, 9 regions, follow up



;

(e) raw data for minimal (circle) and maximal (cross) similarities for 60-day deaths, 9 regions, follow up



;

(f) difference data for minimal (solid) and maximal (dashed) similarities for 60-day deaths, 9 regions, follow up

Figure 4. Minimal and maximal raw and difference trends for combined 9 regions, follow up.

in the very beginning, the mutant variant is located mainly in London [18]. If there is no clear feature change between the two situations, it shows the variant is still gaining momentum and not yet to produce a leading feature. This might also provide a good indication that the vaccines for other COVID-19 viruses should be working for this mutant variant.

4.2. Follow-up results

From Figure 4, we observe that feature similarities fluctuates around 0. This indicates the leading features are pretty independent of each other. In some sense, the underlying behaviour of B.1.1.7 is versatile and less predictable. Combining the preliminary results and the follow-up ones, we shall reach a conclusion that B.1.1.7 has no clear feature and might be very easy in adopting new environments.

5. Conclusions and future work

In this study, we devise 3-valued features to track the dynamical behaviours of SARS-CoV-2 VOC 202012/01. There features serve as the feature database for matching. By preliminary study and follow-up research, our study shows there is no clear leading features for B1.1.7. It also indicates that the virus is hard to pin down its properties and is versatile and hard to predict. The advantages for this study on the features are

a. It is a complete set features and serve as a matching database;
b. It could couple with machine learning and big data analysis to train and locate the features;
c. It could server as the pilot study to understand the potential features of the virus.

There are also some disadvantages:

a. As the length of each datum vector increases, the complexity of computation also exponentially increases;
b. It is hard to practically link the 3-valued features with real natural or scientific properties.

As for the future research, we could apply machine learning techniques on searching and selecting the optimal datum dimension. We could even lift the 3-valued features by adding more features into the candidate featured vectors. We could also associate the 3-valued features with other variables to yield a meaning interpretation of the extracted optimal features or associate them with randomness of genetic codes [17]. Another issue is whether a virus contains a leading feature would increase the severity of death rate of COVID-19 [18].

Acknowledgements

This work is supported by the Humanities and Social Science Research Planning Fund Project under the Ministry of Education of China (Grant No. 20XJAGAT001).

Conflict of interest

No potential conflict of interest was reported by the authors.

References

1. European Centre for Disease Prevention and Control, Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom, *Threat Assess. Brief*, **2020** (2020), 1–13.
2. A. S. Lauring, E. B. Hodcroft, Genetic Variants of SARS-CoV-2—What Do They Mean?, *JAMA*, **325** (2021), 529–531.
3. Public Health England, Investigation of novel SARS-CoV-2 variant Variant of Concern 202012/01 Technical briefing 2, 2020.
4. R. S. Baric, Emergence of a Highly Fit SARS-CoV-2 Variant, *N. Engl. J. Med.*, **383** (2020), 2684–2686.
5. J. W. Tang, P. A Tambyah, D. C. Hui, Emergence of a new SARS-CoV-2 variant in the UK, *J. Infect.*, **82** (2020), E27–E28.
6. J. Wise, Covid-19: New coronavirus variant is identified in UK, *BMJ*, **371** (2020), m4857.
7. New coronavirus variants aren't cause for alarm yet, but mutations could make COVID-19 harder to fight, experts say, Accessed at 11th, January, 2021. Available from: https://www.usatoday.com/story/news/health/2021/01/09/new-coronavirus-strains-variants-not-yet-cause-for-more-covid-vaccine-concerns-experts-say/6575267002/.
8. Coronavirus mutation no reason to panic: virologist, Accessed at 10th, January, 2021. Available from: https://www.dw.com/en/coronavirus-mutation-no-reason-to-panic-virologist/a-56008220.
9. *World Health Organization*, SARS-CoV-2 Variants, Accessed at 11th, January, 2021. Available from: https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/.
10. K. Kupferschmidt, Viral mutations may cause another 'very, very bad' COVID-19 wave, scientists warn, *Science*, Jan. 5, 2021.
11. *Centers for Disease Control and Prevention*, Interim: Implications of the Emerging SARS-CoV-2 Variant VOC 202012/01, Accessed at 10th, January, 2021.
12. E. Callaway, Could new COVID variants undermine vaccines? Labs scramble to find out, *Nature*, **589** (2021), 177–178.
13. GOV.UK, Coronavirus (COVID-19) in the UK, Accessed at 9th, January, 2021. Available from: https://coronavirus.data.gov.uk/details/download.
14. I. Jolliffe, Principal component analysis, in *International Encyclopedia of Statistical Science*, Springer, Berlin, 2011.
15. S. Nagarajaiah, B. Basu, Output only modal identification and structural damage detection using time frequency & wavelet techniques, *Earthquake Eng. Eng. Vib.*, **8** (2009), 583–605.
16. GOV. UK, Coronavirus (COVID-19) in the UK, Accessed at 12th, May, 2021. https://api.coronavirus.data.gov.uk/v2/data?areaType=region&metric=newCasesBySpecimenDate&metric=newDeaths28DaysByDeathDate&metric=newDeaths60DaysByDeathDate&format=csv&release=2021-05-12
17. R. M. Chen, Randomness for Nucleotide Sequences of SARS-CoV-2 and Its Related Subfamilies, *Comput. Math. Methods Med.*, **2020** (2020), 1–8.

18. *European Centre for Disease Prevention and Control*, Risk related to spread of new SARS-CoV-2 variants of concern in the EU/EEA-first update, ECDC: Stockholm, 21 January 2021.

Appendix

Appendix A: Time-series raw data for 9 regions and 3 metrics

Table 1. Time-series sorted data for 9 regions and 3 metrics.

date	region	area label	M1	M2	M3
2020/3/27	Yorkshire and The Humber	1	153	20	20
2020/3/28	Yorkshire and The Humber	1	207	23	23
2020/3/29	Yorkshire and The Humber	1	228	29	29
⋮	⋮	⋮	⋮	⋮	⋮
2021/1/6	Yorkshire and The Humber	1	2287	36	39
2021/1/7	Yorkshire and The Humber	1	1508	31	36
2021/1/8	Yorkshire and The Humber	1	281	18	20
2020/3/27	East of England	2	244	40	40
2020/3/28	East of England	2	183	37	37
2020/3/29	East of England	2	208	49	49
⋮	⋮	⋮	⋮	⋮	⋮
2021/1/6	East of England	2	4621	89	90
2021/1/7	East of England	2	3527	79	85
2021/1/8	East of England	2	394	20	22
⋮	⋮	⋮	⋮	⋮	⋮
2020/3/27	London	8	817	126	126
2020/3/28	London	8	577	120	120
2020/3/29	London	8	597	141	142
⋮	⋮	⋮	⋮	⋮	⋮
2021/1/6	London	8	8985	110	122
2021/1/7	London	8	5906	106	110
2021/1/8	London	8	912	40	43
2020/3/27	West Midlands	9	279	54	54
2020/3/28	West Midlands	9	323	71	71
2020/3/29	West Midlands	9	366	71	71
⋮	⋮	⋮	⋮	⋮	⋮
2021/1/6	West Midlands	9	4758	71	78
2021/1/7	West Midlands	9	2232	64	69
2021/1/8	West Midlands	9	271	20	24

Appendix B: Difference datum vectors

(1) Difference datum vector for cases: (the length of each R_i^k is 287)

- $R_1^1 = (54, 21, 57, 34, \cdots, -742, -386, -779, -1227)$
- $R_2^1 = (-61, 25, 92, 22, \cdots, -2407, -2193, -1094, -3133)$
- $\vdots$
- $R_8^1 = (-240, 20, 271, 66, \cdots, -3310, -4615, -3079, -4994)$
- $R_9^1 = (44, 43, 65, -18, \cdots, -1429, -871, -2526, -1961)$

(2) Difference datum vector for 28-day death:

- $R_1^2 = (3, 6, 1, 7, \cdots, -3, 2, -5, -13)$
- $R_2^2 = (-3, 12, 0, 24, \cdots, -2, -9, -10, -59)$
- $\vdots$
- $R_8^2 = (-6, 21, 10, 27, \cdots, -4, -25, -4, -66)$
- $R_9^2 = (17, 0, 4, 17, \cdots, 0, 6, -7, -44)$

(3) Difference datum vector for 60-day death:

- $R_1^3 = (3, 6, 1, 7, \cdots, -7, 1, -3, -16)$
- $R_2^3 = (-3, 12, 0, 24, \cdots, -5, -12, -5, -63)$
- $\vdots$
- $R_8^3 = (-6, 22, 9, 27, \cdots, -4, -20, -12, -67)$
- $R_9^3 = (17, 0, 4, 17, \cdots, -7, 6, -9, -45)$

Appendix C: Distance tensor product matrices $DTP_{i,j,h}^k$

Table 2. Distance tensor product matrices $DTP^{k}_{i,j,h}$.

	$\vec{f}_1$	$\vec{f}_2$	$\cdots$	$\vec{f}_{2185}$	$\vec{f}_{2186}$	$\cdots$		$\vec{f}_1$	$\vec{f}_2$	$\cdots$	$\vec{f}_{2185}$	$\vec{f}_{2186}$
R^1_{11}	47.96	48.41	$\cdots$	48.41	47.96	$\cdots$	$R^1_{1,41}$	10.26	10.58	$\cdots$	10.58	10.26
R^1_{21}	78.79	79.71	$\cdots$	79.71	78.79	$\cdots$	$R^1_{2,41}$	10.80	10.59	$\cdots$	10.59	10.80
R^1_{31}	38.25	38.71	$\cdots$	38.71	38.25	$\cdots$	$R^1_{3,41}$	12.09	12.26	$\cdots$	12.26	12.09
R^1_{41}	77.42	77.75	$\cdots$	77.75	77.42	$\cdots$	$R^1_{4,41}$	11.24	10.81	$\cdots$	10.81	11.24
R^1_{51}	90.03	91.10	$\cdots$	91.10	90.03	$\cdots$	$R^1_{5,41}$	14.21	14.30	$\cdots$	14.30	14.21
R^1_{61}	54.87	55.26	$\cdots$	55.26	54.87	$\cdots$	$R^1_{6,41}$	11.79	11.96	$\cdots$	11.96	11.79
R^1_{71}	52.00	52.60	$\cdots$	52.60	52.00	$\cdots$	$R^1_{7,41}$	8.92	8.94	$\cdots$	8.94	8.92
R^1_{81}	105.40	106.39	$\cdots$	106.39	105.40	$\cdots$	$R^1_{8,41}$	19.67	19.45	$\cdots$	19.45	19.67
R^1_{91}	70.33	71.01	$\cdots$	71.01	70.33	$\cdots$	$R^1_{9,41}$	11.00	10.26	$\cdots$	10.26	11.00
R^2_{11}	3.45	3.82	$\cdots$	3.82	3.45	$\cdots$	$R^2_{1,41}$	2.84	3.39	$\cdots$	3.39	2.84
R^2_{21}	7.61	7.93	$\cdots$	7.93	7.61	$\cdots$	$R^2_{2,41}$	5.14	5.40	$\cdots$	5.40	5.14
R^2_{31}	3.54	3.65	$\cdots$	3.65	3.54	$\cdots$	$R^2_{3,41}$	5.16	5.29	$\cdots$	5.29	5.16
R^2_{41}	6.95	7.05	$\cdots$	7.05	6.95	$\cdots$	$R^2_{4,41}$	6.08	5.91	$\cdots$	5.91	6.08
R^2_{51}	9.49	9.64	$\cdots$	9.64	9.49	$\cdots$	$R^2_{5,41}$	5.32	5.32	$\cdots$	5.32	5.32
R^2_{61}	4.95	5.01	$\cdots$	5.01	4.95	$\cdots$	$R^2_{6,41}$	3.60	3.91	$\cdots$	3.91	3.60
R^2_{71}	5.07	5.11	$\cdots$	5.11	5.07	$\cdots$	$R^2_{7,41}$	3.39	3.37	$\cdots$	3.37	3.39
R^2_{81}	8.45	8.71	$\cdots$	8.71	8.45	$\cdots$	$R^2_{8,41}$	5.48	5.13	$\cdots$	5.13	5.48
R^2_{91}	6.97	7.06	$\cdots$	7.06	6.97	$\cdots$	$R^2_{9,41}$	6.61	6.07	$\cdots$	6.07	6.61
R^3_{11}	4.72	4.91	$\cdots$	4.91	4.72	$\cdots$	$R^3_{1,41}$	2.84	3.39	$\cdots$	3.39	2.84
R^3_{21}	7.96	8.25	$\cdots$	8.25	7.96	$\cdots$	$R^3_{2,41}$	5.14	5.40	$\cdots$	5.40	5.14
R^3_{31}	3.86	3.99	$\cdots$	3.99	3.86	$\cdots$	$R^3_{3,41}$	5.16	5.29	$\cdots$	5.29	5.16
R^3_{41}	7.18	7.35	$\cdots$	7.35	7.18	$\cdots$	$R^3_{4,41}$	6.08	5.91	$\cdots$	5.91	6.08
R^3_{51}	9.88	10.05	$\cdots$	10.05	9.88	$\cdots$	$R^3_{5,41}$	5.32	5.32	$\cdots$	5.32	5.32
R^3_{61}	4.98	5.05	$\cdots$	5.05	4.98	$\cdots$	$R^3_{6,41}$	3.47	3.78	$\cdots$	3.78	3.47
R^3_{71}	5.14	5.20	$\cdots$	5.20	5.14	$\cdots$	$R^3_{7,41}$	3.39	3.37	$\cdots$	3.37	3.39
R^3_{81}	8.46	8.75	$\cdots$	8.75	8.46	$\cdots$	$R^3_{8,41}$	5.52	5.18	$\cdots$	5.18	5.52
R^3_{91}	7.23	7.32	$\cdots$	7.32	7.23	$\cdots$	$R^3_{9,41}$	6.61	6.07	$\cdots$	6.07	6.61

Appendix D: List of $DTP^k_{j,h}$

Table 3. $DTP^1_{j,h}$ (top block), $DTP^2_{j,h}$ (middle block), $DTP^3_{j,h}$ (bottom block).

sub-vectors	$h=1$	$h=2$	$h=3$	$\cdots$	$h=2184$	$h=2185$	$h=2186$
$j=1$	615.04	620.94	616.95	$\cdots$	616.95	620.94	615.04
$j=2$	17.57	605.54	589.53	$\cdots$	589.53	605.54	617.57
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$j=40$	122.02	122.34	121.67	$\cdots$	121.67	122.34	122.02
$j=41$	109.98	109.15	109.18	$\cdots$	109.18	109.15	109.98
$j=1$	56.47	57.97	56.70	$\cdots$	56.70	57.97	56.47
$j=2$	45.67	45.53	45.15	$\cdots$	45.15	45.53	45.67
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$j=40$	57.48	57.57	56.99	$\cdots$	56.99	57.57	57.48
$j=41$	43.63	43.80	44.03	$\cdots$	44.03	43.80	43.63
$j=1$	59.39	60.87	59.59	$\cdots$	59.59	60.87	59.39
$j=2$	48.22	48.15	47.86	$\cdots$	47.86	48.15	48.22
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$j=40$	57.52	57.60	57.02	$\cdots$	57.02	57.60	57.52
$j=41$	43.54	43.72	43.96	$\cdots$	43.96	43.72	43.54

Appendix E: Optimal rejection distances and features

Table 4. Min/max rejection distances and features: cases (top block), 28-day deaths (middle block) and 60-day deaths (bottom block).

batch	$j=1$	$j=2$	$j=3$	$\cdots$	$j=39$	$j=40$	$j=41$
min^1_j	348.79	382.57	303.97	$\cdots$	94.98	99.11	97.85
$argmin^1_j$	$\vec{f}_{324}$	$\vec{f}_{342}$	$\vec{f}_{685}$	$\cdots$	$\vec{f}_{452}$	$\vec{f}_{445}$	$\vec{f}_{1013}$
max^1_j	621.16	620.42	477.92	$\cdots$	112.05	122.57	114.38
$argmax^1_j$	$\vec{f}_{496}$	$\vec{f}_{112}$	$\vec{f}_{536}$	$\cdots$	$\vec{f}_{65}$	$\vec{f}_{342}$	$\vec{f}_{578}$
min^2_j	44.07	40.88	40.74	$\cdots$	47.60	51.45	41.48
$argmin^2_j$	$\vec{f}_{1093}$	$\vec{f}_{1046}$	$\vec{f}_{302}$	$\cdots$	$\vec{f}_{1092}$	$\vec{f}_{585}$	$\vec{f}_{814}$
max^2_j	58.10	45.98	45.87	$\cdots$	54.89	57.74	47.76
$argmax^2_j$	$\vec{f}_{974}$	$\vec{f}_{998}$	$\vec{f}_{732}$	$\cdots$	$\vec{f}_{851}$	$\vec{f}_{41}$	$\vec{f}_{224}$
min^3_j	47.13	41.33	42.91	$\cdots$	47.41	51.45	41.40
$argmin^3_j$	$\vec{f}_{1093}$	$\vec{f}_{1046}$	$\vec{f}_{660}$	$\cdots$	$\vec{f}_{1092}$	$\vec{f}_{585}$	$\vec{f}_{814}$
max^3_j	60.88	48.36	47.07	$\cdots$	54.90	57.77	47.72
$argmax^3_j$	$\vec{f}_{974}$	$\vec{f}_{730}$	$\vec{f}_{732}$	$\cdots$	$\vec{f}_{851}$	$\vec{f}_{731}$	$\vec{f}_{224}$