A Context-based Time Series Analysis and Prediction Method for Public Health Data

Asako Uraki,[1] Yasushi Kiyoki,[1] Koji Murakami,[2] Akira Kano[3]

[1] Keio University, Keio Research Institute at SFC, Tokyo, Japan
aco@sfc.keio.ac.jp, kiyoki@sfc.keio.ac.jp
[2] Prevent Science Co., Ltd, Tokyo, Japan
murakami@preventme.co.jp
[3] Fujimino Emergency Hospital, Saitama, Japan
kano@koyu-kai.jp

The important process of time series analysis for public health data is to determine target data as a semantic discrete value, according to a context from continuous phenomenon around our circumstance. The phenomenon is expressed as continuous values or discrete data along with time, independent of context. Difference of situations in the phenomenon on time axis expresses one of the key features of time series data, and differences are reflected with adjacent discrete values. Typically, each field of experts has their own fields' specific and practical knowledge to specify an appropriate target part of data which contains the key features of their intended context in each analysis. Those are often implicit, thus not defined as systematically and quantitatively. In this paper, we present a context-based time series analysis and prediction method for public health data. The most essential point of our approach is to express a basis of time series context as the combination of the following 5 elements (1: granularity setting on time axis, 2: feature extraction method, 3: time-window setting, 4: differential computing function, and 5: pivot setting) to determine target data as semantic discrete values, according to the time series context of analysis for public health data.

Keywords:
context-based system,
differential computing,
time-series analysis,
time-series prediction,
public health data,
big data,
AI,
cyber-physical system

University of Maribor Press

DOI https://doi.org/10.18690/um.feri.5.2023.1
ISBN 978-961-286-745-4

1 Introduction

Analysis and Prevention for future situation from the past and current data are important activities to realize preemptive medicine for human health and early detection of spreading infection disease in nature and societies. As our background researches, our semantic computing method [9,10,12] and 5D World Map system [1,2,3,4,5,6,7,8,11] are applied to analysis from the viewpoint of personal health-situation and spreading of infection disease. The important process of time series analysis for public health data is to determine target data as a semantic discrete value, according to a context from continuous phenomenon around our circumstance. The phenomenon is expressed as continuous values or as just a raw discrete data along the time, independent of context. Difference of situations in the phenomenon on time axis expresses one of the key features of time series data, and differences are reflected with adjacent discrete values.

Typically, each field of experts has their own field's specific and practical knowledge to specify an appropriate target part of data which contains the key features of their intended time series context for each analysis. However, those are often implicit therefore not defined systematically and quantitatively.

In this paper, we present a context-based time series analysis and prediction method for public health data. The most essential point of our approach is to express a basis of time series context as the combination of the following 5 elements (1: granularity setting on time axis, 2: feature extraction method, 3: time-window setting, 4: differential computing function, and 5: pivot setting) to determine target data as the semantic discrete value according to the time series context of analysis and prediction for public health data. As our experiment, we realized analysis and prediction by applying actual public health data.

Ordinally in time-series data analysis and prediction, we determine the target part of data along with time axis with an implicit expert knowledge as each context. Our main contribution is to make it possible to explicitly express each context for determination of the target part of data. Then we are able to express certain context quantitatively, to make compatibility, to share the context quantitatively among the different analysis and prediction environment. By changing the context, we can get the different results for discussion and comparison between the different point of view, and to be able to review analytical results of phenomena among users and analyst. Our system enables to express expert knowledge of analysis and prediction on each specific field, and this makes it possible to analyze and discuss interdisciplinary between different fields. The meaning of our new context definition is to fix the closed world of the semantic differential computing on time axis from viewpoint of database system. So, we can normalize the targe part of data according to a context, and it means we can compare the feature extracted from the target data.

One of the main features of our method is to get different results by switching the time series context. Our new concept to express time series context by 5 elements is shown in Figure 1.

If we apply our new context definition to prediction, we are able to realize comparison and discussion between several prediction results for the same data by switching the time series context. Therefore, we are not focusing to evaluate the output results with the real data. Our method realizes quantitative comparison between different time series contexts.

The context definition is quite important to deal with interdisciplinary phenomenon between human and providence of nature, such as the field of health, medication, environment, and culture which are having time axis. As our experiment, we realized our system and context definition in the field of public health to analyze and prediction of infection disease. The experiments show the prediction feasibility of our method in the field of public health data, effectiveness to generate results for discussion regarding switching context, and applicability to express time series context of an expert knowledge for analysis and prediction as combination of the 5 elements to make the knowledge explicit and quantitative expression.

In the next section, we present our method by explaining overview and definitions of data and functions. In section 3, we show two experiments by applying actual public health data, and we conclude this paper in section 4.

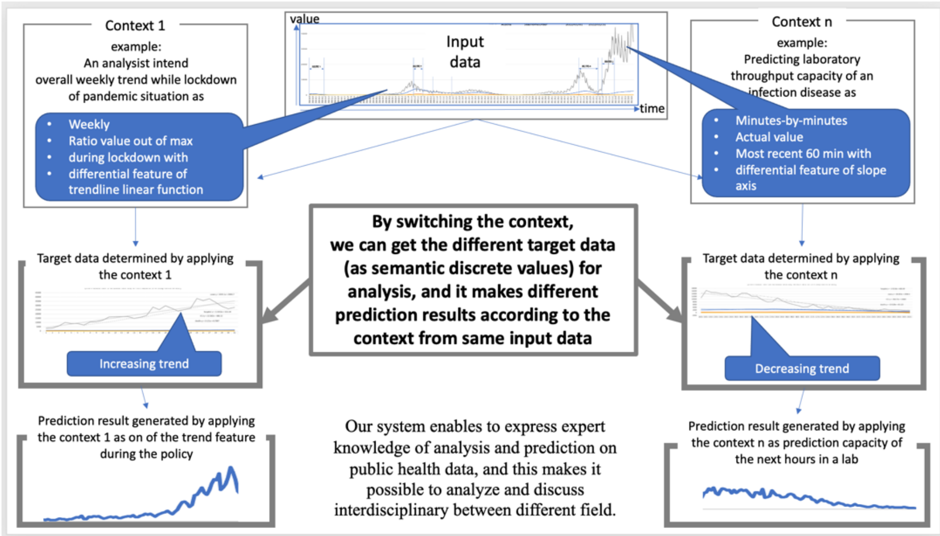

Figure 1: Concept of the context-based time series analysis and prediction method for public health data. The left side shows context 1 (ex: for overall weekly trend while lockdown of pandemic situation), and the right side another context n (ex: prediction laboratory throughput capacity of an infection disease). By switching those contexts, we can get the different target data (as semantic discrete values) for analysis, and it makes different prediction results according to the context from same input data.

Source: own.

2 A Context-based Time Series Analysis and Prediction Method for Public Health Data

In this section, we define a context-based time series analysis and prediction method for public health data. First, we explain overview of our method to process the 5 elements of context to analysis and prediction for public health data. Then we define 5 elements to express basis of a time series context to generate target data as semantic discrete values of analysis and prediction according to the time series context, and we also define of input data, target data, and output data of our method.

2.1 Overview of the Context-based Time Series Analysis and Prediction Method

We realize our method by the following 4 steps to get analysis and prediction result. Overview of the proposing method is shown in Figure 2.

Step 1: Input data (continuous data/raw discrete data) along with time (before considering time series context)

Input time series data for reference (IRD)
Input time series data for prediction (ITD)

Step2: Define a time series context for analysis and prediction by combination of the 5 elements

Granularity setting on time axis (GS)
Feature extraction method (FEM)
Time-window setting (TWS)
Differential computing function (DCF)
Pivot setting on time axis for prediction (PV)

Step3: Extract target data and pivot according to the 5 elements defined in Step2

Confirmed target data (CRD)
Time point of Pivot on ITD (PV)

Step4: Output prediction result according to the context

Output prediction data (OPD)

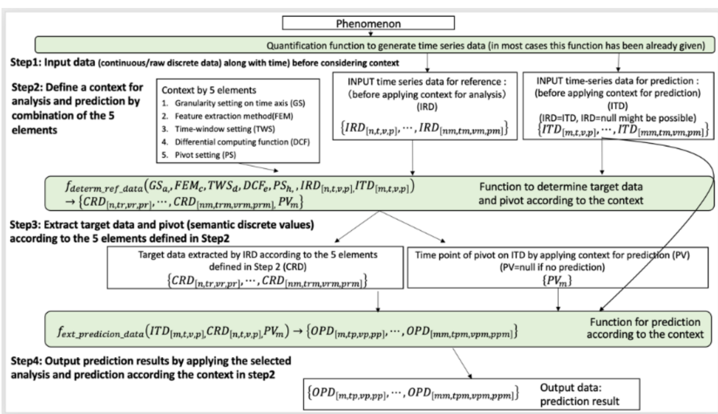

Figure 2: Overview of a context-based time series analysis and prediction method for public health data. We realize our model by the 4 steps to output prediction result.

Source: own.

2.2 Data Structure and Definition of 5 elements to express a time series context to determine semantic discrete values as target of analysis and prediction

Our approach is to express the basis of time series context for analysis and prediction as combination of the following 5 elements, Granularity setting on time axis (GS), Feature extraction method (FEM), Time-window setting (TWS), Differential computing function (DCF), and Pivot setting on time axis for prediction (PV) to determine the appropriate target data as semantic discrete values on time axis.

In the field of analysis of time series data, we have many viewpoints for expressing time series context to fix target data not only those elements but also for the specific criteria on each field. In this paper, we only focused on the analysis and prediction of public health data, and we designed it by applying above 5 elements with the expert knowledge which are previously implicit in the brain of each analyst. Therefore, by applying the above 5 elements, we are able to express a time series context to fix target data and pivot on time axis for analysis and prediction in the field of public health data. Each element has each role to fix the target data as explaining in the following subsections.

Granularity setting on time axis (GS): Setting granularity on time axis corresponds to fix a semantic sampling rate of target data as semantic discrete values. We have many conceptual numeral systems on time axis and its sampling size according to a context. The granularity setting in time (GS) includes two variables, original granularity in time (OG), and target granularity in time (TG). We define data structure of GS as the following.

In general, as we can feel in daily life, we have many kinds of numeral systems on time axis, such as base 24 system (24 ticks of conceptual and hierarchal structures between day and hour), sexagesimal system (60 ticks of conceptual and hierarchal structures for second, minute, and hour). We also have several conceptual systems on time axis, such as hierarchal text writing system of paper (hierarchal system between phrase, sentence, section, and paper), music hierarchal system (hierarchal system between note, phrase, section, movement, and a piece of music) Moreover, we use specific segmentation to fix an important part of data by applying additional information other than data itself, such as heart beat rate during workout or not, blood pressure value in duration of medication or not. We have introduced the conceptual tick and the hierarchal system of music data as "grain and tree-structured granularity" for music analysis in our previous research [14]. By switching the system and selecting the level of granularity, we can specify an appropriate grain of data according to a context. Ordinally in the field of data analysis, the granularity of input data will be used as is the granularity of the targe data. Our model makes it possible to explicit desired granularity of context to generate target data for each analysis and prediction. If we have an attention on larger granularity, the difference between data expresses comprehensive feature. Other cases, if we have another attention on smaller granularity, the difference between data expresses detailed differential feature of the data.

$$GS_a \supset \{OG_b, TG_b\} \tag{1}$$
$$(a = 1, am)(a = granularity\ control\ method\ id, am = maximum\ number\ of\ granularity\ control\ method)$$
$$(b = 1, bm)(b = granulairty\ id, bm = maximum\ number\ of\ granulairty)$$

Feature extraction method (FEM): Setting of the feature extraction method means normalization of values of target data in a closed world of context for analysis and prediction. We define data structure of FEM as the following. The feature extraction method is the quantification method to generate value of the target data on time axis, such as ratio out of maximum number, specific quantity, semantic feature value on specific viewpoint, and so on.

$$FEM_c \tag{2}$$
$$(c = 1, cm)(c = feature\ extraction\ method\ id, cm = maximum\ number\ of\ feature\ extraction\ method)$$

Time-window setting (TWS): Setting the time- window means selection of intended range of time axis for each context. Time-window setting (TWS) contains two variables, starting time point (TS) and ending time point (TE). We define data structure of TWS as the following.

$$TWS_d \supset \{TS_{tp}, TE_{tp}\} \tag{3}$$
$$(d = 1, dm)$$
$$(d = time\_window\ setting\ id, dm = maximum\ number\ of\ time\_window\ setting\ method)$$
$$(tp = time\ point\ on\ IRD)$$

Differential computing function (DCF): Setting of the differential computing method means switching type of ruler to calculate differential feature between adjacent data on time axis for each context. We define data structure of the DCF as the following.

The differential computing method is a quantification method to calculate value of the difference of the target data on time axis, such as regressive curve, tilt/angle, slope, trend line linear, substruction, ratio between each substruction, a color system to calculate distance between colors, and so on. The determination of the differential computing method is often considered with the feature extraction method since in many cases, as both two are developed together in specific research field such as metadata generation method and the specific calculation system for the metadata.

$$DCF_e \tag{4}$$
$$(e = 1, em)$$
$$(e = differential\ computing\ \text{function}\ id, em = maximum\ number\ of\ differential\ computing\ \text{function})$$

Pivot setting (PV): Setting pivot means to fix an effective timing of starting prediction by applying target data of each context. The pivot is a starting time point of predication that matched the conceptual meaning of the starting time of target data, such as the same day of week, same month in a year, having a similarity in supportive data, having a similarity in target data, and so on. We define data structure of PV as the following.

$$PV_h \tag{5}$$
$$(h = 1, hm)$$
$$(h = \text{pivot setting}\ id, hm = maximum\ number\ of\ pivot\ setting)$$
$$(tp = time\ point\ on\ ITD)$$

2.3 Data Structure and Definition of Input Data, Target data, and Output Data of the Context-based Time Series Analysis and Prediction Method

Our input data consists of the following two data, 1) Input Reference Data (IRD) and 2) Input Target Data (ITD). IRD is a base data for picking up target data according to a context. ITD is a prediction target data by applying the target data also according to the context. We define the data structure of the two inputs data as the following formula. ITD=IRD might be possible when if the IRD is also desired prediction data, and ITD=null might be also possible when if only analysis. We define IRD and ITD as the followings.

Input time series data for reference (IRD):

$$\left\{IRD_{[n,t,v,p]}, \cdots, IRD_{[nm,tm,vm,pm]}\right\} \tag{6}$$
$$(n = 1, nm)(nm = maximum\ number\ of\ IRD)$$
$$(t = 1, tm)(tm = maximum\ number\ of\ time\ point)$$
$$(v = 1, vm)(vm = maximum\ value)$$
$$(p = 1, pm)(pm = maximum\ number\ of\ parameter)$$

Input time series data for prediction (ITD):

$$\left\{ITD_{[m,t,v,p]}, \cdots, ITD_{[mm,tm,vm,pm]}\right\} \tag{7}$$
$$(m = 1, mm)(mm = maximum\ number\ of\ ITD)$$

Confirmed target data (CRD):
The data structure of the target data is formalized as the following.

$$\left\{CRD_{[n,tr,vr,pr]}, \cdots, CRD_{[nm,trm,vrm,prm]}\right\} \tag{8}$$
$$(rt = 1, trm)(trm = maximum\ number\ of\ time\ point)$$
$$(vr = 1, vrm)(vrm = maximum\ value)$$
$$(pr = 1, prm)(prm = maximum\ number\ of\ parameter)$$

Output prediction data (OPD):

Data structure of the output data structure is formalized as the following.

$$\{OPD_{[m,tp,vp,pp]},\cdots,OPD_{[mm,\ tpm,vpm,ppm]}\} \quad (9)$$
$$(tp = 1, tpm)(tpm = maximum\ number\ of\ time\ point)$$
$$(vp = 1, vpm)(vpm = maximum\ value)$$
$$(pp = 1, ppm)(ppm = maximum\ number\ of\ parameter)$$

2.4 Functions of the Context-based Time Series Analysis and Prediction Method

Function to determine target data and pivot:

We define the function to determine target data and pivot as the following.

$$f_{determine\_ref\_data}(GS_a, FEM_b, FW_{[ts,te]}, DCM_c, PV_h, IRD_{[n,t,v,p]}, ITD_{[nm,t,v,p]}) \rightarrow \{CRD_{[n,tr,vr,pr]},\cdots,CRD_{[nm,trm,vrm,prm]}, PV_m\} \quad (10)$$

Function for prediction according to the context:

We define the function for prediction as the following.

$$f_{extract\_predicion\_data}(ITD_{[m,t,v,p]}, CRD_{[n,t,v,p]}, PV_m) \rightarrow \{OPD_{[m,tp,vp,pp]},\cdots,OPD_{[mm,tpm,vpm,ppm]}\} \quad (11)$$

3 Experiment Study

We realized our experimental studies by applying COVID-19 number of confirmed cases data as the actual phenomenon of one of the public health issues.

Experiment 1:

This experiment is to predict Covid-19 number of cases to expect tightness of testing throughput capacity of the Covid-19 (infection disease) in a laboratory during the coming week in Austria with single parameterized input data. The point of this experiment is to know what day of the week we need to expect maximum number of processing. A testing laboratory has implicit statistics that the number of confirmed cases is mostly synchronized with tightness day of the testing throughput capacity. To determine reference data for prediction, we set the following two contexts by an expert of analysis in the field of public health data. The time series context 1 is to reflect the most recent situation for prediction. The context 2 is to reflect the most similar situation in for prediction. By applying the two contexts to determine reference data to the prediction, we expect to get different prediction result according to each context.

The important knowledge in the context setting of experiment 1 is to focus on the number of confirmed cases on Sundays, since most of the testing sites are closed on Sundays in Austria by business restrictions same as other DACH countries, and the confirmed cases on Sundays are expressing only the results from the regional core hospitals to care relatively severe conditioned patients. The expert of analysis in the field of public health data who set those contexts assumes that ratio comparing by the Sunday's number expresses one of the key features of situation of the following days on the week. Therefore, starting time point of the TWS and the PV is on Sunday in this experiment.

Input data and time series context 1 for experiment 1: to predict tightness of testing throughput capacity of the Covid-19 (infection disease) in a laboratory during the coming week in Austria **by reflecting the most recent situation for prediction.**

Input time series data for reference (IRD) = Covid-19 number of confirmed cases in Austria which is published by ECDC [16]

Input time series data for prediction (ITD) = prediction data is same as IRD, Covid-19 number of confirmed cases in Austria

Granularity setting (GS)
Original granularity (OG) = daily
Target granularity (TG) = daily
Feature extraction method (FEM) = actual number of confirmed cases
Time-window setting (TWS) = most recent 1 week starting from Sunday on IRD
Differential computing function (DCF) = ratio between starting point number of cases
Pivot setting on ITD (PV) = the most recent Sunday on ITD

The confirmed reference data (CRD) for context 1 of experiment1 and the prediction output (OPD) are shown in Figure 3.

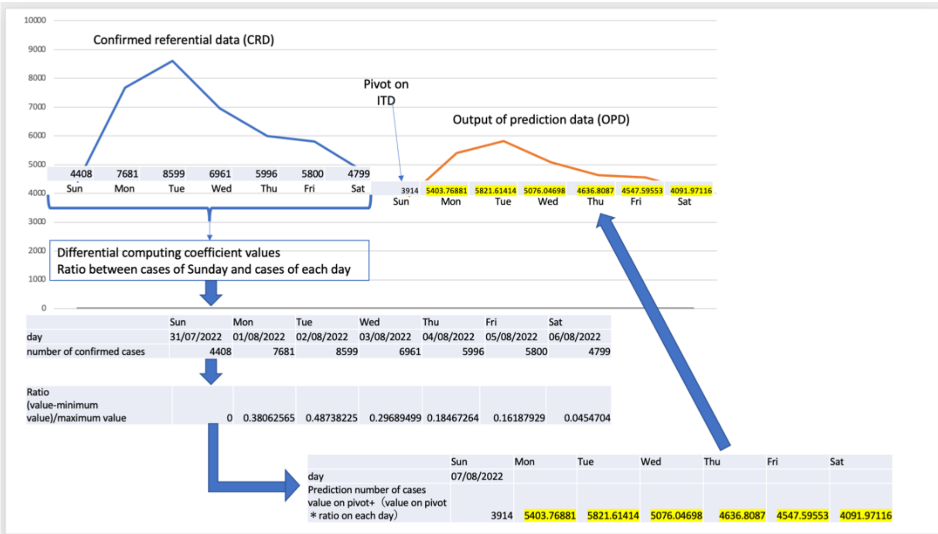

Figure 3: The confirmed reference data (CRD) and the prediction output (OPD) for context 1 of experiment1. The top left data (blue line) shows the most recent one-week data starting from Sunday which reflect the context 1, and the number of confirmed cases on each day calculated as the ratio between cases of Sunday and cases of each day as shown in the table. The ratio applied to predict the next one-week number of cases as the right bottom table. The output prediction data (OPD) is shown in top right (orange line).

Source: own.

Input data and time series context 2 for experiment 1: to predict tightness of testing throughput capacity of the Covid-19 (infection disease) in a laboratory during the coming week in Austria **<u>by reflecting the most similar past situation (amount of change and absolute value) for prediction</u>**

Input time series data for reference (IRD) = Covid-19 number of confirmed cases in Austria which is published by ECDC [16]

Input time series data for prediction (ITD) = prediction data is same as IRD, Covid-19 number of confirmed cases in Austria

Granularity setting on time axis (GS)

Original granularity (OG) = daily (for reference data), weekly (for determination of reference data)

Target granularity (TG) = daily for prediction
Feature extraction method (FEM) = actual number of confirmed cases

Time-window setting (TWS) = most recent 1 week starting from Sunday that matched condition (condition: combination of slope trend (over two weeks of decreasing trend) and absolute value (the time point right after under 5000 cases))

Differential computing function (DCF) = ratio between starting point number of cases

Pivot setting on ITD (PV) = the most recent Sunday on ITD

The Process to determine the confirmed reference data (CRD) for the context 2 of experiment1 is shown in Figure 4, and the confirmed reference data (CRD) for context 1 of experiment1 and the prediction output (OPD) are shown in Figure 5.

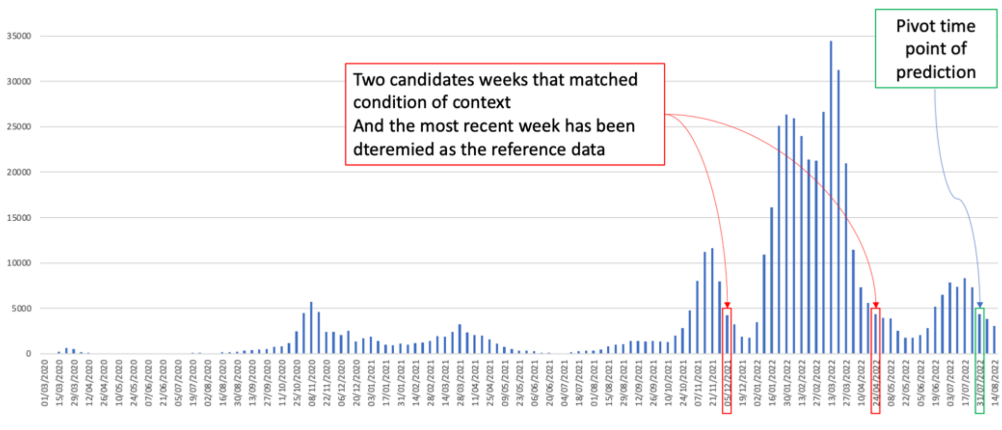

Figure 4: Process to determine the confirmed reference data (CRD) for the context 2 of experiment1. The figure shows number of confirmed cases on every Sundays (as weekly data), and the red squares show two candidate weeks that matched condition of time-window setting (TWS) of the context 2. The most recent week from the candidates (the right red square) has been determined as the reference data (CRD). The green square shows time point of the pivot (PV) for prediction.

Source: own.

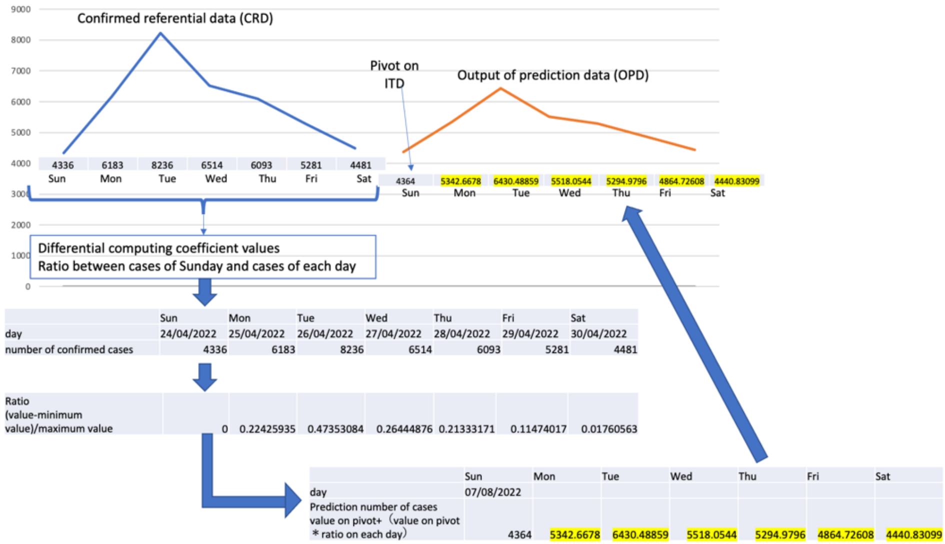

Figure 5: The confirmed reference data (CRD) and the prediction output (OPD) for context 2 of experiment1. The top left data (blue line) shows the one-week data that matched condition starting from Sunday which reflect the context 2, and the number of confirmed cases on each day calculated as the ratio between cases of Sunday and cases of each day as shown in the table. The ratio applied to predict the next one-week number of cases as the right bottom table. The output prediction data (OPD) is shown in top right (orange line).

Source: own.

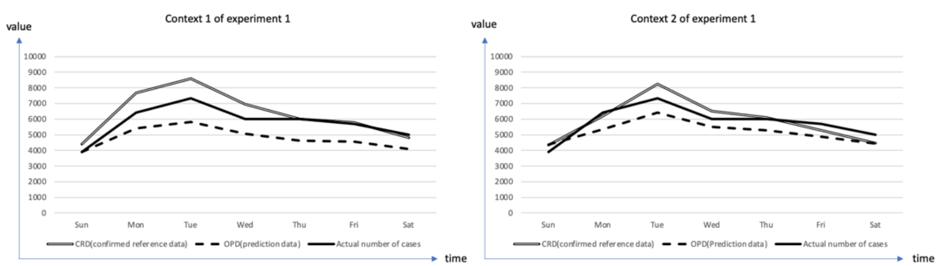

Figure 6: Comparison between CRD (confirmed referential data), OPD (output prediction data), and the actual number of confirmed cases. Left-side chart is result of context 1 of experiment1(<u>by reflecting the most recent situation for prediction</u>), and the right-side chart is result of context 2 of experiment (<u>by reflecting the most similar past situation (amount of change and absolute value) for prediction</u>). Dash line shows prediction data, double line shows CRD (confirmed referential data), and the solid line shows actual number of cases which later published from ECDC. By comparing OPD and the actual number of cases, the context 2 is closer than context 1.

Source: own.

Comparison between CRD (confirmed referential data), OPD (output prediction data), and the actual number of confirmed cases are shown in Figure 6. The left-side chart is result of context 1 of experiment1, and the right-side chart is result of context 2 of experiment. Dash line shows prediction data, double line shows CRD (confirmed referential data), and the solid line shows actual number of cases which later published from ECDC. By comparing OPD and the actual number of cases, the context 2 is closer than context 1. Results of the experiment 1 show following discussions.

Prediction feasibility of our method in the field of public health data
Realized quantitative comparison between different time series context

Effectiveness to generate results for discussion regarding switching the setting of 5 elements to reflect better settings of time series context to the other prediction quantitatively

Applicability to express time series context of an expert knowledge for analysis and prediction as the combination of 5 elements, to make the knowledge explicit and quantitative expression

Experiment 2:

Experiment 2 is focusing on switching two major variant, Alpha and Delta, during pandemic situation of Covid-19 in 2022 with multiparametric input data. This experiment is to predict the timing to reach over 90% ratio of a spreading (increasing) variant while switching with another variant. To determine reference data for prediction, we set the following two contexts to select input time series data for reference (IRD), by an expert of analysis in the field of public health data. The time series context 1 is to reflect the closer population, and the time series context 2 is to reflect the closer population density. By applying those contexts, we selected Covid-19 confirmed number of variant cases of nationwide data of United Kingdom and city level data of London as the input time series data for reference (IRD) which is published by government of United Kingdom [15]. Those areas have already switched majority from Alpha variant to Delta variant completely, and the Delta variant reached over 90% ratio. And we also selected Covid-19 confirmed number of variant cases of Saitama as input time series data for prediction (ITD) which are still in the half time point of switching majority from Alpha variant to Delta variant.

The important knowledge in the context setting of experiment 2 is to focus on closer population and closer population density. The expert of analysis in the field of public health data who set those contexts assumes that rapidness of switching two major variant is corresponding population density, and rapidness is slower in higher density area and rapidness is quicker in lower density area. We expect to get different prediction result according to each context, and we also expect get basis to compare it for discussion of the expert's assumption.

In this experiment, we extract rapidness of switching two major variants by applying the rapidness calculation function which we already introduced in [13] as the differential computing function (DCF), and we reflect it to predict rapidness of Saitama (ITD) for the second half situation. The concept of the experiment 2 is shown in the Figure 7. For the other time series context were also set by the expert from an epidemiological laboratory in Japan for public health data analysis to express their desired context for this comparison.

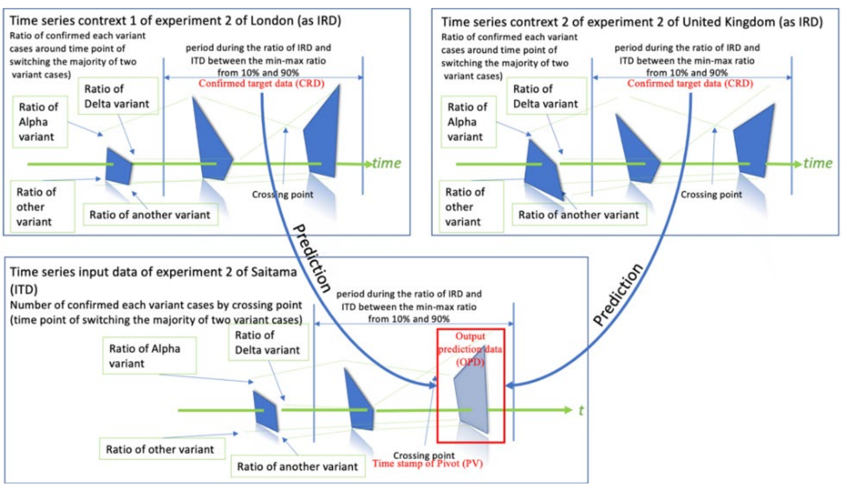

Figure 7.: Concept of the experiment2 to predict the timing to reach over 90% ratio of a spreading (increasing) variant while switching with another variant. The top left shows context 1 by applying London data (IRD) and target data (CRD), and the top right shows context 2 by applying United Kingdom data (IRD) and target data (CRD). The bottom shows time series prediction data of Saitama (ITD).

Source: own.

Input data and time series context 1 for experiment 2: to predict the timing to reach over 90% ratio of a spreading (increasing) variant while switching with another variant in Saitama **by reflecting London (closer feature is population density) for prediction**

Input time series data for reference (IRD) = Covid-19 number of confirmed Alpha and Delta variant cases out of all confirmed variant cases in London (already switched the majority of the two variants) which is published by government of United Kingdom [15]. Population density of London is 4761 people/square kilometers.

Input time series data for prediction (ITD) = Covid-19 number of confirmed Alpha and Delta variant cases out of all confirmed variant cases in Saitama (still in the half time point of switching majority from Alpha variant to Delta variant) which is collected by hospital and testing facility in Saitama. Population density of Saitama is 6127 people/square kilometers.

Granularity setting on time axis (GS)
Original granularity (OG) = weekly
Target granularity (TG) = weekly
Feature extraction method (FEM) = ratio of confirmed each variant case out of all confirmed variant cases
Time-window setting (TWS) = period during the ratio of IRD and ITD between the min-max ratio from 10% and 90%
Differential computing function (DCF) = area size calculation function [13], ratio calculation function of area-a size between IRD and ITD, switching rapidness coefficient table between two variants
Pivot setting on ITD (PV) = crossing point (time point of switching the majority of two variant cases) on ITD
The input time series data for reference (IRD) of London and the time series input data for prediction (ITD) of Saitama is shown in Figure 8.

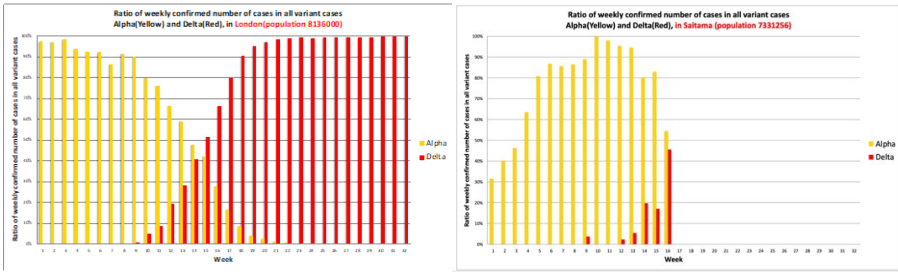

Figure 8: The left chart shows time series input data for reference (IRD) of London. The right chart shows time series input data for prediction (ITD) of Saitama. This data is time-series ratio of confirmed patient of each variant of Covid-19 which expresses situation of majority of variant of each area. London data shows situation is that this area is already switched the majority of the two variants. Saitama data shows situation that this area is still in the half time point of switching majority from Alpha variant to Delta variant.
Source: own.

Determined target data for analysis (CRD) for context 1 of the experiment 2 is shown in Figure 9. The size of the area A, B, C and D express rapidness of the switching and the smaller size shows quicker switching and larger size shows slower switching.

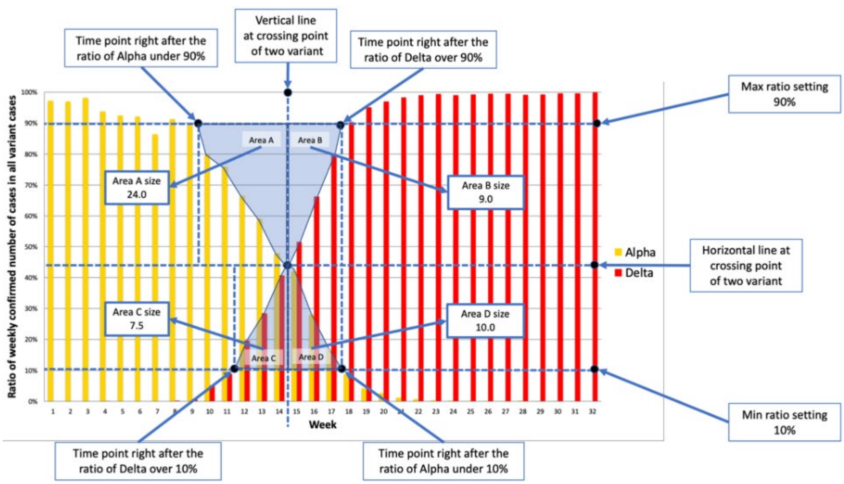

Figure 9: Target data for analysis (CRD) of London for experiment 2 by applying 5 elements of context which expressing rapidness of switching two major variants. The size of the area A, B, C and D express rapidness of the switching and the smaller size shows quicker switching and larger size shows slower switching.
Source: own.

Input data and time series context 2 for experiment 2: to predict the timing to reach over 90% ratio of a spreading (increasing) variant while switching with another variant in Sweden **by reflecting United Kingdom (population density is not closer) for prediction**

Input time series data for reference (IRD) = Covid-19 number of confirmed Alpha and Delta variant cases out of all confirmed variant cases in United Kingdom (already switched the majority of the two variants) which is published by government of United Kingdom [15]. Population density of United Kingdom is 257 people/square kilometers (not closer).

Input time series data for prediction (ITD) = Covid-19 number of confirmed Alpha and Delta variant cases out of all confirmed variant cases in Saitama (still in the half time point of switching majority from Alpha variant to Delta variant) which is collected by hospital and testing facility in Saitama. Population density of Saitama is 6127 people/square kilometers (not closer).

Granularity setting on time axis (GS)
Original granularity (OG) = weekly
Target granularity (TG) = weekly
Feature extraction method (FEM) = ratio of confirmed each variant case out of all confirmed variant cases
Time-window setting (TWS) = period during the ratio of IRD and ITD between the min-max ratio from 10% and 90%
Differential computing function (DCF) = area size calculation function [13], ratio calculation function of area-a size between IRD and ITD, switching rapidness coefficient table between two variants
Pivot setting on ITD (PV) = crossing point (time point of switching the majority of two variant cases) on ITD

The input time series data for reference (IRD) of United Kingdom and the time series input data for prediction (ITD) of Saitama is shown in Figure 10.

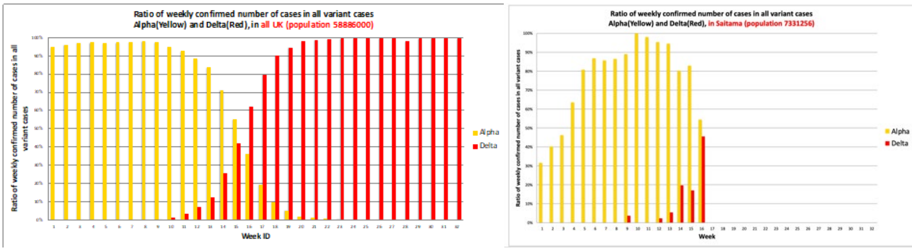

Figure 10: The left chart shows time series input data for reference (IRD) of United Kingdom. The right chart shows time series input data for prediction (ITD) of Saitama. This data is time-series ratio of confirmed patient of each variant of Covid-19 which expresses situation of majority of variant of each area. United Kingdom data shows situation is that this area is already switched the majority of the two variants. Saitama data shows situation that this area is still in the half time point of switching majority from Alpha variant to Delta variant.

Source: own.

Determined target data for analysis (CRD) for context 1 of the experiment 2 is shown in Figure 11. The size of the area A, B, C and D express rapidness of the switching and the smaller size shows quicker switching and larger size shows slower switching. The area size analysis of the input time series data for prediction (ITD) of Saitama is shown in Figure 12.

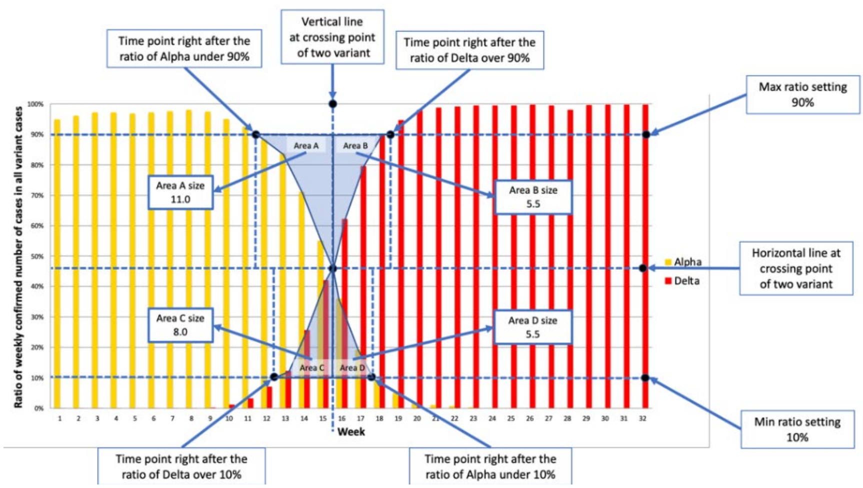

Figure 11: Target data for analysis (CRD) of All UK for experiment 2 by applying 5 elements of context which expressing rapidness of switching two major variants. The size of the area A, B, C and D express rapidness of the switching and the smaller size shows quicker switching and larger size shows slower switching.

Source: own.

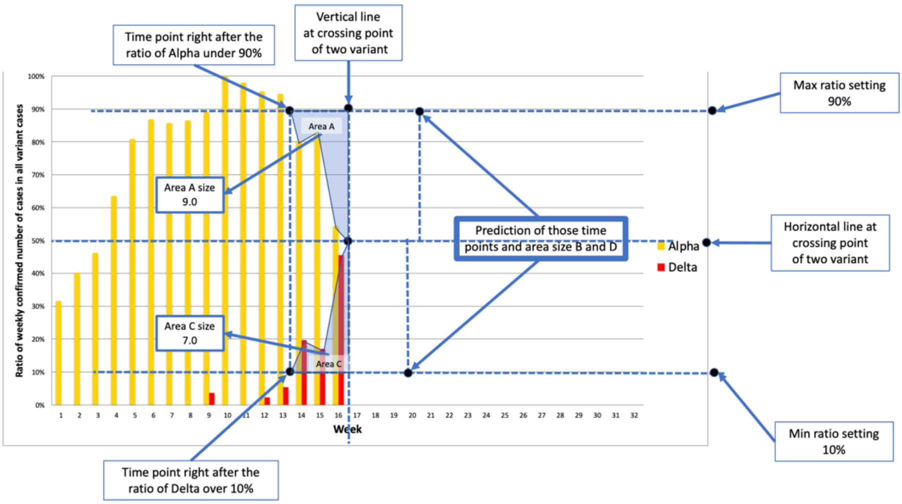

Figure 12: Area size analysis of the input time series data for prediction (ITD) of Saitama. The area size of A expresses switching force of Delta variant against Alpha variant. The area size of C expresses endurance force of Alpha variant against Delta variant. This area size is differential computing result and it reflects feature of the Saitama data for the prediction in the next step processing.

Source: own.

Output data and comparison between context 1 and 2 of experiment 2:

By applying the London data and United Kingdom data of rapidness ratio between area A, B, C and D, we can get the prediction result of the Saitama after the crossing point as area size and the timing of the data over 90% ratio. The output prediction data (OPD) of the experiment 2 is shown in Figure 13 for both context 1 and 2.

Results of the experiment 2 shows following discussions.

- Prediction feasibility of our method in the field of public health data with the multiparametric input data
- Realized quantitative comparison between different time series context on different places which have different environmental feature
- Effectiveness for discussion regarding
- switching the setting of 5 elements with field specific condition
- processing different kind of granularity on time axis

- to reflect better settings of time series context to the other prediction quantitatively
- Applicability of field specific function of public health data analysis to our presenting method

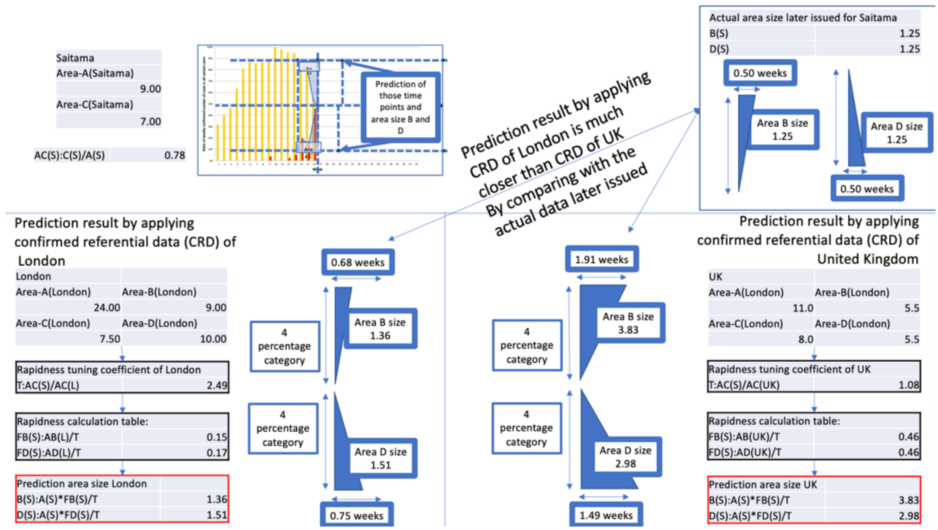

Figure 13: Result of Experiment-2 as the output prediction value (OPD) of Saitama. The bottom left values show prediction result by applying London data, and the bottom right values show prediction result by applying UK data. The top right values show the actual area size and week which later issued for Saitama. By comparing those results, the prediction results by applying CRD of London is much closer than CRD of UK to the actual data.

Source: own.

4 Conclusion

We have presented a context-based time series analysis and prediction method for public health data. The most essential point of our approach is to express a basis of context as the combination of the following 5 elements (1: granularity setting on time axis, 2: feature extraction method, 3: time-window setting, 4: differential computing function, and 5: pivot setting) to determine target data as the semantic discrete value according to the context of analysis for public health data. As our experiment, we realized analysis and prediction by applying public health data. As our future work, we will design appropriate evaluation in this field to express the essence of our method, we will apply our method not only for prediction but also for datamining/analysis/search, and we will extend our method and the system to realize mutual understanding and knowledge sharing on global human-health issues in the world-wide scope.

Acknowledgement

We are grateful to Dr. Shiori Sasaki for essential and helpful discussion of this study.

References

[1] Yasushi Kiyoki, Xing Chen, “A Semantic Associative Computation Method for Automatic Decorative-Multimedia Creation with “Kansei” Information” (Invited Paper), The Sixth Asia-Pacific Conferences on Conceptual Modelling (APCCM 2009), 9 pages, January 20-23, 2009.

[2] Yasushi Kiyoki, Xing Chen, Shiori Sasaki, Chawan Koopipat, “A Globally-Integrated Environmental Analysis and Visualization System with Multi-Spectral & Semantic Computing in “Multi-Dimensional World Map””, Information Modelling and Knowledge Bases XXVIII, pp.106-122,2017

[3] Yasushi Kiyoki and Saeko Ishihara: “A Semantic Search Space Integration Method for Meta-level Knowledge Acquisition from Heterogeneous Databases,” Information Modeling and Knowledge Bases (IOS Press), Vol. 14, pp.86-103, May 2002.

[4] Yasushi Kiyoki, Shiori Sasaki, Nhung Nguyen Trang, Nguyen Thi Ngoc Diep, "Cross-cultural Multimedia Computing with Impression-based Semantic Spaces," Conceptual Modelling and Its Theoretical Foundations, Lecture Notes in Computer Science, Springer, pp.316-328, March 2012.

[5] Yasushi Kiyoki: “A “Kansei: Multimedia Computing System for Environmental Analysis and Cross-Cultural Communication,” 7th IEEE International Conference on Semantic Computing, keynote speech, Sept. 2013.

[6] Shiori Sasaki, Yusuke Takahashi, Yasushi Kiyoki: “The 4D World Map System with Semantic and Spatiotemporal Analyzers,” Information Modelling and Knowledge Bases, Vol.XXI, IOS Press, 18 pages, 2010.

[7] Totok Suhardijanto, Yasushi Kiyoki, Ali Ridho Barakbah: “A Term-based Cross-Cultural Computing System for Cultural Semantics Analysis with Phonological-Semantic Vector Spaces,” Information Modelling and Knowledge Bases XXIII, pp.20-38, IOS Press, 2012.

[8] Chalisa Veesommai, Yasushi Kiyoki, Shiori Sasaki and Petchporn Chawakitchareon, "Wide-Area River-Water Quality Analysis and Visualization with 5D World Map System", Information Modelling and Knowledge Bases, Vol. XXVII, pp.31-41, 2016.

[9] Chalisa Veesommai, Yasushi Kiyoki, “Spatial Dynamics of The Global Water Quality Analysis System with Semantic-Ordering Functions”. Information Modelling and Knowledge Bases, Vol. XXIX, 2018.

[10] Yasushi Kiyoki, Asako Uraki, Chalisa Veesommai, “A Seawater-Quality Analysis Semantic-Space in Hawaii-Islands with Multi-Dimensional World Map System”, 18th International Electronics Symposium (IES2016), Bali, Indonesia, September 29-30, 2016.

[11] Shiori Sasaki and Yasushi Kiyoki, "Real-time Sensing, Processing and Actuation Functions of 5D World Map System: A Collaborative Knowledge Sharing System for Environmental Analysis", Information Modelling and Knowledge Bases, Vol. XXVIII, IOS Press, pp. 220-239, May 2016.

[12] Shiori Sasaki, Koji Murakami, Yasushi Kiyoki, Asako Uraki: “Global & Geographical Mapping and Visualization Method for Personal/Collective Health Data with 5D World Map System,” Information Modelling and Knowledge Bases (IOS Press), Vol. XXXII, pp. 134 – 149, 2020.

[13] Yasushi Kiyoki, Koji Murakami, Shiori Sasaki, Asako Uraki, "Human-Health-Analysis Semantic Computing & 5D World Map System" Information Modelling and Knowledge Bases (IOS Press), Vol. XXXIII, pp. 141 – 151, 2022.
[14] A. Ijichi and Y. Kiyoki: "A Kansei Metadata Generation Method for Music Data Dealing with Dramatic Interpretation "Information Modeling and Knowledge Bases, IOS Press, Vol.XVI, pp.170-182,(2004).
[15] UK Health Security Agency. Variant of Concern Technical Briefing 23. Available at: https://www.gov.uk/government/publications/investigation-of-sars-cov-2-variants-technical-briefings
[16] ECDC European Centre for Disease Prevention and Control, Data on the daily number of new reported COVID-19 cases and deaths by EU/EEA country, Available at: https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country