

Emerging risk forecast system using associative index mining analysis

Joo-Chang Kim¹ · Kyungyong Chung²

Received: 20 June 2016 / Revised: 13 October 2016 / Accepted: 21 November 2016 / Published online: 3 December 2016
© Springer Science+Business Media New York 2016

Abstract Accessibility to the information increased by the development of IT converging technology and therefore various information such as economy, society, technology, culture and etc are distributed without limit to time and place. Data mining that discovers useful information by analyzing the data that accumulated to a large amount is receiving spotlight. By the distribution of IoT machine, the collection of life log has been eased and the interest on personalized service fit to the used is increasing. Companies and nationals use various analysis methods to develop big data that is accumulated for the safety of its people and asset into simple index. Meteorological Administration analyze the weather data and provide service through internet and app by developing weather index analyzing the extent of influence of weather on life, health and industry. The weather index measure by each region by using data collected real time at the weather observation point. Weather index is classified into the life weather index, health weather index and industry weather index by fields and provide coping measures per steps along with precautions. The weather index provided at this present is a simple service that did not consider the cor-

relation of the index. Also it has a problem of not considering the situation by the location of the user. This study suggests emerging risk forecast system using associative index mining analysis. The suggesting method is the monitoring system focused on the user by analyzing the associative index using data mining and forecasting emerging risk by reflecting the user situation information. Index data collects the weather index XML file by using the Rest call method through the Meteorological Administration. The collected weather index analyzed the association relationship and interrelationship between the index by using the Apriori algorithm and Pearson correlation coefficient of the data mining. User situation information collects the surrounding temperature and humidity using the Zigbee communication. Collected temperature and humidity is applied to the daily weather index and health weather index and calculated fit for the situation of the user and the industry weather index is calculated focused on the user by using the association relationship and interrelationship between the index. By using the calculated weather index focused on the user, emerging risk of the life, health and industry scope is forecasted. Emerging risk is classified into life emerging risk, health emerging risk, industry emerging risk by the forecast method and is divided into four steps of low, general, high and danger. Emerging risk forecast system using the associative index mining analysis suggested provides precaution followed in the emerging risk and provides service such as weather index, weather information, user situation information etc of the Meteorological Administration. The suggested emerging risk forecasting system is the user based monitoring simulation that calculates the weather index real-time.

✉ Kyungyong Chung
kyungyong.chung@gmail.com

Joo-Chang Kim
kjc2232@naver.com

¹ Intelligent System Lab., Dept. of Computer Information Engineering, Sangji University, 83, Sangjida-gil, Wonju-si, Gangwon-do 26339, Korea

² School of Computer Information Engineering, Sangji University, 83, Sangjida-gil, Wonju-si, Gangwon-do 26339, Korea

Keywords Weather index · Associative index · Data mining · Emerging risk · Risk forecast

1 Introduction

Users share information through Facebook, Instagram, Twitter, internet community and etc massively produce contents of various fields of economy, society, technology, culture and etc. and is receiving attention as the principal of the new information producer. By this, government established the environment to produce new contents by approaching the wanted information by anyone by opening the public data to the people. Also, collection of various information occurred in the daily life became possible by development of IoT machines such as smart watch, smart band, IP CAM and etc. that have grafted IT convergence technology to the product by the development of converging technology and the interest on the personalized service by the situation of the user is increasing [1,2]. User is using and sharing information real time through the internet and data is accumulated as life log [3] by the usage of information. Various studies to provide differentiated service to the user by analyzing and using life log that is accumulated massively is lively being proceeded. Data mining that discovered meaningful and new information in the data that is accumulated massively, is receiving spotlight. Various index is being developed for safety of the people by using data such as disaster, calamity, weather and etc among the massive data that is accumulated in the information society [4]. Representative index related to this are Disaster Risk Index (DRI), Natural Disaster Risk Index (NDRI) and weather index [5]. Disaster Risk Index is the index developed to reduce damage of human life and assets by representative natural disasters such as earthquake, typhoon, flood and drought and is the danger level is classified into seven steps [6]. Natural disaster index is showed into the index the extent of influence of the natural disaster to human by considering date rate per million people and the occurrence number for 30 years of natural disaster such as earthquake, eruption, tsunami, typhoon, flood, landslide, heat wave etc [7–11].

Meteorological Administration developed weather index by analyzing the extent of influence to the life, health, industry by the weather by analyzing the weather data with close relationship to the national life [1,8,12,13]. Weather index is classified into life weather index, health index and industry weather index by the fields. This is by measuring the weather index per region using collected data real time at the weather observation point of the region. Weather index provides measures and precautions per step through internet webpage and application. However, the weather index of Meteorological Administration that is in service now has a problem of not considering the variable by personal situation as it is the regional based service [14,15].

To solve such problem, this study suggests emerging risk forecast system using the associative index mining analysis. The suggested method is reflecting the association of index

and situation of the user to the weather index calculation and is the user based emerging risk service that recalculates based on the user. By using the weather index based on the user, it forecasts emerging risk that occur to the user and provide measurement and precautions by the emerging risk step. Data for the associative index mining analysis is composited of transaction collecting weather index XML through public data portal [16–18]. Collected data is analyzed using the Apriori algorithm of mining and *Pearson* correlation coefficient. Emerging risk is classified into 4 steps of low, general, high, and danger. This is classified into life emerging risk, health emerging risk, industry emerging risk by the forecast method.

This study is organized as follows. Section 2 describes relevant studies of disasters risk index. Section 3 represents the emerging risks forecast using the associative index mining analysis. Section 4 describes the system and implementation, and Sect. 5 offers a conclusion.

2 Related disasters risk index

Many companies and nations are providing by expressing into index the influence to the safety of the people and to the development of the industry by data analysis of disaster, calamity and human life damage. There services are providing by NRI, Mortality Risk Index (MRI), NDRI, DRI, Disaster Deficit Index (DDI), NDH, DRI, NHIM, and PVI [9,13,19–21].

DRI is the index presented in the United Nations Development Programme to prepare for the disaster damage and danger. This compares the death danger extent of each nation by using the Emergency Events Database statistics data and measures the danger level of population exposed to drought, flood, typhoon and earthquake among natural disaster. DRI is classified into seven steps from 0 to 7 and the danger of death is low as it is closer to 0 and the danger of death is high as it is closer to 7. DRI has the limit of not considering natural disaster of small or medium scale that is occurred frequently as it only targets big disaster that has big scale of damage. Also, data collection of human life damage is inaccurate and there is hard weakness to apply to the working level [6]. MRI is a similar method with DRI measuring by considering the annual average number of the death forecast by the danger factor and danger exposure per nation by landslide, flood, typhoon and earthquake. Considering danger factors are disaster frequency, strength and number of human damage etc. There could be error in case of country with frequent drought as the data related to drought is not considered.

RRI measures the natural disaster danger decrease extent of each nation as it is the project that is implemented in Costa Rica, Guatemala, Nicaragua, assassination Bay, Honduras, Panama and Dominican Republic for natural disaster

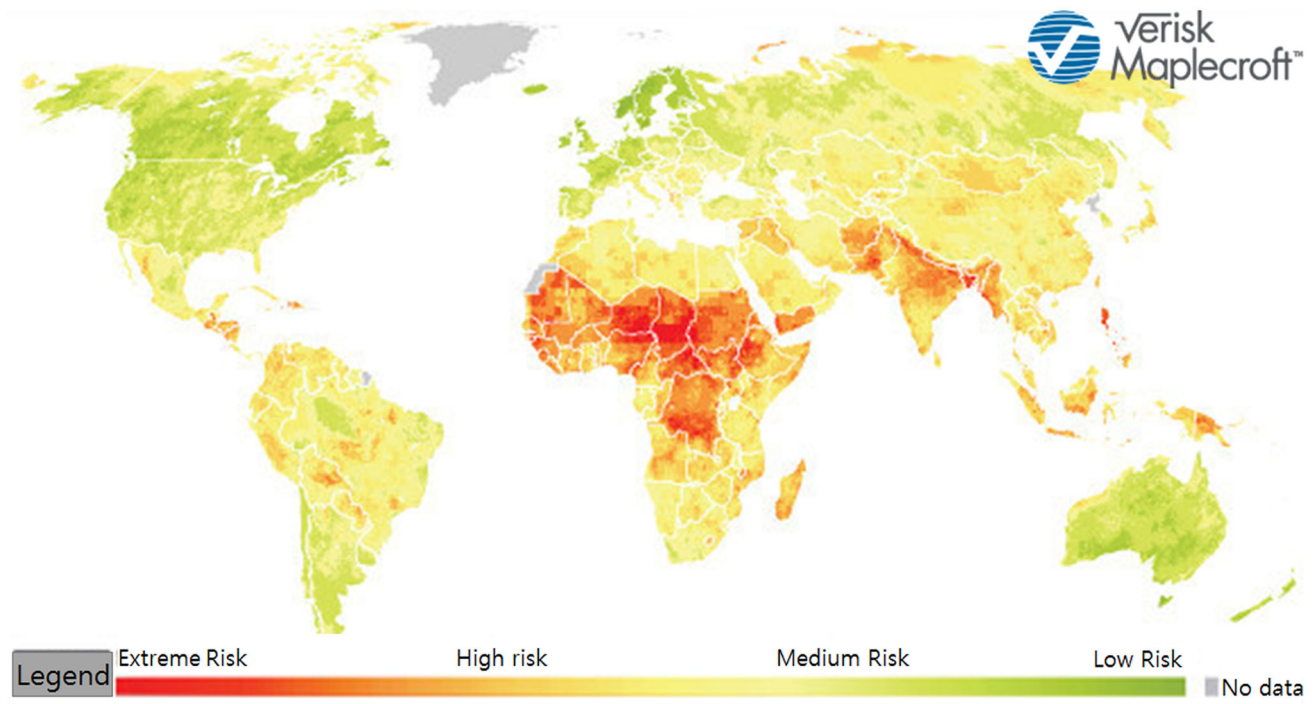


Fig. 1 Climate change vulnerability index service by Maplecroft

reduction measures. This includes exposure to danger factor, human and asset damage, national asset and environment management, calamity measures and etc reducing danger by analyzing the factors of the disaster. Danger factor means environment and natural resources, social economy, means of livelihood, land and environment building and etc. Survey is implemented to the professionals with professional knowledge related with disaster related danger of the concerning region for the index measurement. NHIM (Natural Hazard Index for Megacities) is the index measuring the damage extent on the natural disaster targeted to 50 cities worldwide. Earthquake, storm and flood is determined as the main danger factor and volcano, forest fire and freeze etc. is classified into the secondary danger factor. NHIM considers danger degree of natural disaster, human calamity, social calamity and financial calamity.

CCVI represents the vulnerability of populations according to changing the extreme climate related factors and environment parameters. It combines the risk of exposure with the country's capacity to adapt to climate change. It is shown in seven steps of low risk, medium risk, high risk, extreme risk by Maplecroft¹. Figure 1 shows the climate change vulnerability index service by Maplecroft [7].

PVI measures direct or indirect damage of natural disasters. This is expressed in an index using the exposure extent of natural disaster, extent of social economic weakness, extent

of natural disaster elasticity etc. There is human and city growth rate, population density, ratio of the poor, GDP and etc in the natural disaster exposure extent. There is human poverty rate, aging rate, inequality, unemployment rate, price and etc in the social economic weakness. Natural disaster elasticity measures coping skill on crisis. DDI measures on the national economic response competence on large scale disaster by judgement. It is the index measuring the extent of competence that government can solve the maximum forecast of economic loss cause of natural disaster. It means that loss amount will show large compared to the financial competence when DDI is larger than 1.

NDH is the index made by analyzing the danger extent of earthquake, volcano, landslides, flood, drought etc. of certain region and measures by analyzing the correlation between the risk of death and economic loss and disaster risk for each nation. This includes the occurrence percentage of disaster, population and asset exposed to disaster, population exposed to certain disaster and weakness of asset etc., but has the limit that it does not consider weakness to natural disaster. Earthquake Disaster Risk Index (EDRI) is the index comparing disaster danger related with earthquake and measure the extent that earthquake related factors influence to the overall danger. As the factors of EDRI, there are earthquake intensity, earthquake range, earthquake frequency, population exposure, exposed facilities, emergency response, and recovery capabilities, etc. [19].

¹ Maplecroft, www.maplecroft.com.

3 Emerging risks forecast using associative index mining analysis

3.1 Data collection and preprocessing

As the data for calculating the weather index by situation of the user and analyze the association between the index, it collects the Meteorological Administration weather index, user temperature and user humidity [14]. The weather index data collection used the OPEN API provided by the public data portal and uses the certification key provided by the portal in the REST signal method and collects the weather index XML. REST signal method can collect XML file by inserting URL composited by forming the signal URL by compositing the request address, index code, region code and certification key (service key) [16]. Certification key cannot be disclosed outside for information security reason therefore is shown in random letters. Weather index XML includes date, index information, index code and signal date of today, tomorrow and day after tomorrow in DOM form [14, 17]. It is formed of tree form easy to collect and extract information and has high program transplanted. Weather index XML file collects data of June, 2016 and forms the collected data from the same region and time and forms one transaction. Industry weather index has stopped service since January 2016 therefore the weather information provided by the Meteorological Administration through the separate weather index monitoring program to collect by using the weather information of the observation point [13].

Weather index code provided by the Meteorological Administration is C01, C02, C03, C04 and etc granting index from A to Z for the transaction formation as it is not continuous. The life weather index and health weather index from the collected weather index is in four steps where step 1 is low, step 2 is general, step 3 high, and step 4 very high. Indus-

try weather index is in three steps where step 1 is bad, step 2 general and step 3 good [14]. As the steps and meaning per step of the weather index is different, the index value is normalized into four steps for analysis. Index weather index is normalized to 4 for very bad, 3 for bad, 2 for general, and 1 for good. Weather index that went through normalization process is expressed in (index) + (index value) and showed as A2 when the food poisoning index is general. Table 1 shows the normalization process of weather index expression.

Weather index transaction is formed of csv form that is easy to express the weather index among the ‘arff’ form and ‘csv’ form which are the representative data forms. Sensory temperature index (F), frostbite index (G), freeze index (H), atmospheric diffusion index (I), heating energy index (T), and pollen index (Y) has different providing term from Meteorological Administration and therefore is excluded from the mining analysis. Second transaction of Table 2 which is T002 means that A3 has a high food poisoning index, B3 has a high putrefaction index. Transaction ID (TID) is formed of T001 to T180. Table 2 shows the weather index transaction.

User situation information for emerging risk forecast is collected real-time the surrounding temperature and humidity by the GPS location of the user by Zigbee communication [10, 15, 21]. Zigbee communication is the communication that is possible to use in short distance and has low energy consumption amount and the communication distance is extended to maximum of 100 m when installing external antenna which is gives it the advantage of making it possible to simply realize from the system. Zigbee module used for the user data collection has temperature, humidity, illumination, infrared light sensor installed basically and various information collection is possible such as fine dust, body temperature, air pressure, ozone, and etc when installing external module [13, 22]. Modules for Zigbee communication are formed of receiver and transmitter. The surrounding temperature and

Table 1 Normalization process of weather index expression

Weather index	ID	Weather index	ID
Food Poisoning Index	A	Masonry Work Index	N
Putrefaction Index	B	Finish Work Index	O
Discomfort Index	C	Carriage Index	P
Heat Index	D	Fishery Products Index	Q
Ultraviolet Index	E	Ice Cakes Index	R
Sensory Temperature Index	F	Fruits Vegetables Index	S
Frostbite Index	G	Heating Energy Index	T
Freeze Index	H	Cooling Energy Index	U
Atmospheric Diffusion Index	I	Influenza Index	V
Agricultural Facility Index	J	Stroke Index	W
Crop-Dusting Index	K	Asthma Index	X
Foundation Work Index	L	Pollen Index	Y
Skeleton Work Index	M	Skin Problem Index	Z

Table 2 Weather index transaction

TID	A	B	C	D	E	J	K	L	M	N	O	–	Z
T001	A2	B2	C2	D2	E2	J2	K2	L2	M2	N2	O3	–	Z1
T002	A3	B3	C3	D3	E3	J3	K2	L1	M1	N2	O2	–	Z2
T003	A1	B1	C1	D1	E1	J1	K1	L2	M2	N2	O3	–	Z2
T004	A1	B1	C2	D2	E2	J2	K1	L2	M1	N2	O2	–	Z2
T005	A1	B2	C2	D3	E3	J2	K2	L1	M1	N1	O3	–	Z2
T006	A2	B3	C2	D3	E3	J1	K1	L2	M2	N2	O2	–	Z2
T007	A2	B2	C2	D2	E2	J2	K2	L2	M2	N2	O2	–	Z1
T008	A3	B3	C3	D3	E3	J3	K2	L1	M1	N2	O2	–	Z2
T009	A1	B1	C1	D1	E1	J1	K1	L2	M2	N2	O2	–	Z2
T010	A1	B1	C2	D2	E2	J2	K1	L2	M2	N3	O1	–	Z2
T011	A1	B1	C2	D2	E2	J2	K1	L2	M2	N3	O1	–	Z2
T012	A4	B4	C4	D3	E3	J3	K2	L1	M1	N1	O2	–	Z3
T013	A3	B3	C3	D4	E3	J3	K2	L1	M1	N2	O2	–	Z2
T014	A2	B3	C4	D4	E3	J2	K1	L2	M2	N1	O2	–	Z1
T015	A1	B1	C2	D2	E2	J2	K1	L2	M2	N3	O2	–	Z2
T016	A1	B1	C2	D2	E2	J2	K1	L2	M2	N3	O2	–	Z2
–	–	–	–	–	–	–	–	–	–	–	–	–	–
T180	A2	B3	C2	D3	E4	J2	K2	L3	M3	N3	O3	–	Z2

humidity of the user measured at the transmitter is sent to the receiver and the receiver receives the user information and applies to the monitoring system.

3.2 Associative relation analysis of weather index using mining

The associative relative analysis of the weather index uses Apriori algorithm [23–25] and *Pearson* correlation coefficient [8, 10, 26]. Associative relation between the index using algorithm is extracted and the correlation between the index is analyzed by using the *Pearson* correlation. Apriori algorithm is the big data analysis method extracting associative relation by repeating the process of forming the candidate set as the item that is higher than the minimum support by scanning the transaction and determine the minimum support. Weather index associative relation analysis uses the TANAGRA 1.4.50 Version which is the data mining analysis tool [27]. Apriori algorithm analysis brings out different analysis result by setting minimum support, lift, confidence, maximum itemset and when the lift value is 1 or more it is evaluated as a useful rule and the minimum lift value is set as 1.0 and when the confidence value is 0.75 or more, it is evaluated as the useful rule and the minimum confidence value is set to 0.75 [24]. In case the weather index transaction has set item of 4 or more, the maximum set item is set as 4 as it does not satisfy other options. In case the minimum support is 0.5 or more, there are two rules and analyze by reducing the minimum support per 0.01 from 0.5.

Table 3 Apriori algorithm mining analysis result by the minimum support

No.	Min. sup.	Lift	Confidence	Rules
1	0.50	1.522546	95.42975	16
2	0.49	1.688677	94.48826	98
3	0.48	1.841656	95.44644	705
4	0.47	1.837119	95.34907	713
5	0.46	1.835622	95.32872	717
6	0.45	1.831409	95.31174	725
7	0.44	1.829806	95.25784	733
8	0.43	1.825462	95.11731	759
9	0.42	1.825007	94.83604	805
10	0.41	1.831726	94.68015	867

As the result of Apriori algorithm, the lift and confidence showed to be excellent in all section while the minimum support value drastically increases its number of rules from 98 to 705 when it is 0.48 or less, the minimum support is to be set 0.49. Table 3 shows Apriori algorithm mining analysis result by the minimum support.

Result of weather index analysis using Apriori algorithm is shown on Table 4. TID 1 of Table 4, if the antecedent is ‘Q2 = Q2’ and consequent is ‘R2 = R2’, it means that is the present weather index Q is step 2 (general), it means that weather index R will show to be step 2 (general). Table 4 shows the weather index analysis result when the minimum support is 0.49.

Table 4 Weather index mining analysis result when the minimum support is 0.49

TID	Antecedent	Consequent	Lift	Support	Confidence
001	“Q2 = Q2”	“R2 = R2”	1.99	49.72	99.69
002	“Q2 = Q2”	“W2 = W2”–“R2 = R2”	1.99	49.72	98.89
003	“R2 = R2”	“W2 = W2”–“Q2 = Q2”	1.92	49.72	96.74
004	“W2 = W2”–“Q2 = Q2”	“R2 = R2”	1.92	49.72	98.89
005	“W2 = W2”–“R2 = R2”	“Q2 = Q2”	1.92	49.72	98.89
006	“R2 = R2”	“Q2 = Q2”	1.92	49.72	96.74
007	“O2 = O2”–“L2 = L2”	“W2 = W2”–“M2 = M2”	1.92	49.16	95.65
008	“M2 = M2”–“R2 = R2”	“O2 = O2”–“L2 = L2”	1.92	49.16	98.88
009	“O2 = O2”–“L2 = L2”	“M2 = M2”–“R2 = R2”	1.92	49.16	95.65
010	“W2 = W2”–“M2 = M2”	“O2 = O2”–“L2 = L2”	1.92	49.16	98.88
011	“O2 = O2”–“R2 = R2”	“L2 = L2”	1.90	49.72	98.89
012	“L2 = L2”	“O2 = O2”–“R2 = R2”	1.90	49.72	95.70
013	“O2 = O2”–“M2 = M2”–“R2 = R2”	“L2 = L2”	1.90	49.16	98.88
014	“M2 = M2”–“R2 = R2”	“L2 = L2”	1.90	49.16	98.88
–	–	–	–	–	–
098	“A2 = A1”	“O2 = O2”	1.17	50.28	96.77

Correlation analysis is the method of looking for rule between variables by analyzing the similarity between the variables. By analyzing what kind of correlation the variables that occur simultaneously has on each other and the extent of correlation is showed in numbers. There is *Pearson* correlation coefficient and *Spearman* rank-order correlation coefficient as the representative coefficients between the two variables [13, 28]. *Pearson* correlation shows the linear relationship between the two variables and the correlation coefficient is shown as p . The correlation calculates by using covariance and has the independence where correlation does not change even though the two variables are regularly changed using the four fundamental arithmetic operations. Correlation coefficient p has the value from -1.0 to $+1.0$. This has the strength and direction and means they have same direction and similar as it gets closer to $+1$. If the correlation between the two variables gets closer to -1 , it means their direction is different and has the opposite tendency. In case the correlation value is 0 , it means that there is no relationship between the two variables. Equation (1) shows the formula of the *Pearson* correlation coefficient. In Eq. (1), p means correlation coefficient, Cov covariance, X and Y variables [28].

$$p(X, Y) = \frac{\sum_{i=1}^{180} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{180} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{180} (Y_i - \bar{Y})^2}} \quad (1)$$

Spearman rank-order correlation coefficient has the advantage to calculate the data expressed as the comparatively simple and ordinal scale as it applied the correlation calculation formula. It is similar with the *Pearson* correlation

coefficient but it does not use the actual data value and calculates by using the ranking of the data value. This has a disadvantage of having decreasing accuracy if there are many data that are the same and this study uses the *Pearson* correlation coefficient as the weather index transaction that is analyzed in this study repetitively shows the equal data [25].


Weather index transaction normalizes the vector to substitute the *Pearson* correlation coefficient formula. The correlation analysis using *Pearson* correlation coefficient shows the same result even though the variable gives the same weight within the transaction and therefore the index value is maintained to be 1 (low), 2 (general), 3 (high), and 4 (very high) and the index granted to the index is excluded from the transaction. Figure 2 shows the vector normalization process of the weather index transaction. Transaction that normalized the vector analyzes the correlation between the index by applying the *Pearson* correlation from the SPSS 24 version [29].

Table 5 shows the correlation between the weather index using the *Pearson* correlation coefficient. In Table 5, each index is shown from A to Z and the ones in red express negative number. The correlation of A and B means 0.87 and means that food poisoning index (A) and putrefaction index (B) have high probability of being similar. The correlation of A and L in Table 5 is -0.30 and it means that food poisoning index (A) and foundation work index (L) have general probability of showing to be opposite.

Figure 3 shows the correlation distribution of the weather index. In the graph the index shows that they are similar as they get closer to 1. This visualizes the correlation between the weather index and it can be checked visibly in Fig. 3.

Fig. 2 Vector normalization process of the weather index transaction

TID	A	B	C	...	Z
T001	A2	B2	C2	...	Z1
T002	A3	B3	C3	...	Z2
T003	A1	B1	C1	...	Z2
T004	A1	B1	C2	...	Z2
T005	A1	B2	C2	...	Z2
T006	A2	B3	C2	...	Z2
T007	A2	B2	C2	...	Z1
T008	A3	B3	C3	...	Z2
T009	A1	B1	C1	...	Z2
T010	A1	B1	C2	...	Z2
T011	A1	B1	C2	...	Z2
T012	A4	B4	C4	...	Z3
T013	A3	B3	C3	...	Z2
T014	A2	B3	C4	...	Z1
T015	A1	B1	C2	...	Z2
T016	A1	B1	C2	...	Z2
...
T180	A2	B3	C2	...	Z2



TID	A	B	C	...	Z
T001	0.5	0.5	0.5	...	0.25
T002	0.75	0.75	0.75	...	0.5
T003	0.25	0.25	0.25	...	0.5
T004	0.25	0.25	0.5	...	0.5
T005	0.25	0.5	0.5	...	0.5
T006	0.5	0.75	0.5	...	0.5
T007	0.5	0.5	0.5	...	0.25
T008	0.75	0.75	0.75	...	0.5
T009	0.25	0.25	0.25	...	0.5
T010	0.25	0.25	0.5	...	0.5
T011	0.25	0.25	0.5	...	0.5
T012	1	1	1	...	0.75
T013	0.75	0.75	0.75	...	0.5
T014	0.5	0.75	1	...	0.25
T015	0.25	0.25	0.5	...	0.5
T016	0.25	0.25	0.5	...	0.5
...
T180	0.5	0.75	0.5	...	0.5

Table 5 Correlation between the weather index using the *Pearson* correlation coefficient

	A	B	C	D	E	J	K	L	M	N	O	P	Q	R	S	U	V	W	X	Z
A	1.00	0.87	0.80	0.71	0.59	0.60	0.64	-0.30	-0.31	-0.22	0.07	0.74	0.60	0.59	0.76	0.71	0.19	0.75	-0.20	0.14
B	0.87	1.00	0.87	0.85	0.70	0.40	0.75	-0.20	-0.19	-0.41	0.29	0.85	0.79	0.75	0.82	0.75	-0.14	0.68	-0.07	0.09
C	0.80	0.87	1.00	0.89	0.70	0.52	0.59	-0.22	-0.22	-0.23	0.20	0.78	0.71	0.65	0.89	0.71	0.04	0.65	0.13	0.12
D	0.71	0.85	0.89	1.00	0.84	0.60	0.67	-0.33	-0.35	-0.28	0.18	0.78	0.82	0.78	0.84	0.84	0.11	0.67	0.22	0.12
E	0.59	0.70	0.70	0.84	1.00	0.69	0.60	-0.40	-0.40	-0.19	0.07	0.61	0.66	0.63	0.67	0.67	0.33	0.55	0.33	0.05
J	0.60	0.40	0.52	0.60	0.69	1.00	0.45	-0.52	-0.59	0.03	-0.22	0.43	0.34	0.34	0.56	0.56	0.75	0.56	0.19	-0.02
K	0.64	0.75	0.59	0.67	0.60	0.45	1.00	-0.17	-0.20	-0.32	0.32	0.82	0.62	0.64	0.74	0.61	-0.13	0.60	-0.22	-0.14
L	-0.30	-0.20	-0.22	-0.33	-0.40	-0.52	-0.17	1.00	0.96	0.49	0.65	-0.06	-0.22	-0.22	-0.13	-0.36	-0.29	-0.15	-0.04	-0.16
M	-0.31	-0.19	-0.22	-0.35	-0.40	-0.59	-0.20	0.96	1.00	0.44	0.65	-0.10	-0.30	-0.30	-0.18	-0.44	-0.37	-0.22	-0.12	-0.21
N	-0.22	-0.41	-0.23	-0.28	-0.19	0.03	-0.32	0.49	0.44	1.00	0.20	-0.26	-0.34	-0.29	-0.03	-0.25	0.45	0.02	0.14	-0.01
O	0.07	0.29	0.20	0.18	0.07	-0.22	0.32	0.65	0.65	0.20	1.00	0.46	0.23	0.24	0.32	0.11	-0.32	0.29	0.01	-0.03
P	0.74	0.85	0.78	0.78	0.61	0.43	0.82	-0.06	-0.10	-0.26	0.46	1.00	0.72	0.68	0.84	0.76	-0.10	0.83	-0.08	-0.07
Q	0.60	0.79	0.71	0.82	0.66	0.34	0.62	-0.22	-0.30	-0.34	0.23	0.72	1.00	0.97	0.72	0.90	-0.06	0.63	0.25	0.38
R	0.59	0.75	0.65	0.78	0.63	0.34	0.64	-0.22	-0.30	-0.29	0.24	0.68	0.97	1.00	0.71	0.89	-0.03	0.61	0.18	0.45
S	0.76	0.82	0.89	0.84	0.67	0.56	0.74	-0.13	-0.18	-0.03	0.32	0.84	0.72	0.71	1.00	0.74	0.09	0.76	0.07	0.09
U	0.71	0.75	0.71	0.84	0.67	0.56	0.61	-0.36	-0.44	-0.25	0.11	0.76	0.90	0.89	0.74	1.00	0.16	0.79	0.13	0.34
V	0.19	-0.14	0.04	0.11	0.33	0.75	-0.13	-0.29	-0.37	0.45	-0.32	-0.10	-0.06	-0.03	0.09	0.16	1.00	0.26	0.41	0.17
W	0.75	0.68	0.65	0.67	0.55	0.56	0.60	-0.15	-0.22	0.02	0.29	0.83	0.63	0.61	0.76	0.79	0.26	1.00	0.05	0.19
X	-0.20	-0.07	0.13	0.22	0.33	0.19	-0.22	-0.04	-0.12	0.14	0.01	-0.08	0.25	0.18	0.07	0.13	0.41	0.05	1.00	0.38
Z	0.14	0.09	0.12	0.12	0.05	-0.02	-0.14	-0.16	-0.21	-0.01	-0.03	-0.07	0.38	0.45	0.09	0.34	0.17	0.19	0.38	1.00

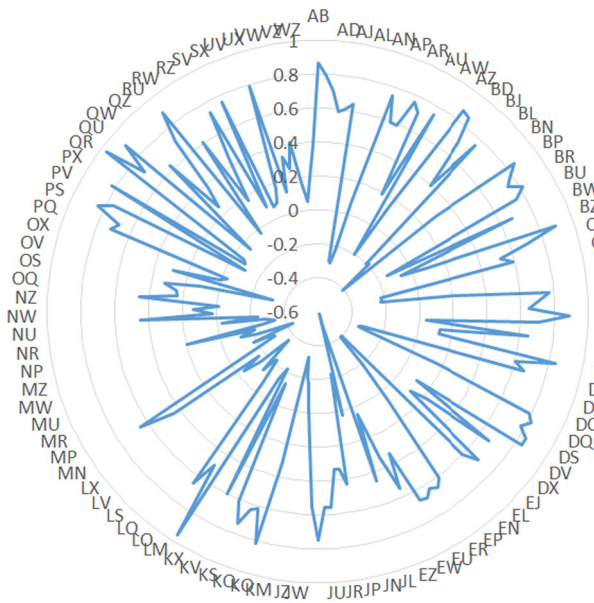


Fig. 3 Correlation distribution of weather index

3.3 Emerging risk forecast according to user situation

Emerging risk by user is the personalized index forecast method by reflecting information such as temperature, humidity, ultraviolet rays, fine dust, wind speed, and etc that changes by the location of the user in the regional weather index service provided by the Meteorological Administration [14]. Life emerging risk [1] and health emerging risk [12] forecasts the emerging risk by applying the tempera-

ture and humidity around the user at the present to the life weather index and health weather index calculation formula that is provided to the Meteorological Administration. Industry emerging risk forecasts the weather index based on the user and forecasts the industry weather index by using the correlation between the weather index correlation and index. There could be difference in case the temperature and humidity around the user at the present is reflected as it is calculated as the outside information of the observation point collected from the Meteorological Administration [14]. To overcome this, weather index is calculated again based on the user. Life weather index and health weather index is calculated real time by reflecting the collected surrounding temperature, humidity of the user using the Zigbee communication. A more accurate life weather index service is possible by the temperature and humidity considering character changing by the situation of the user as it is the core in calculating the life weather index and health weather index [30]. Emerging risk is forecasted by using the average of life weather index and health weather index that is re-calculated. Figure 4 shows the emerging risk forecast process using mining engine.

Emerging risk index uses the correlation of index and the weather index association rule and forecast. Industry weather index uses the maximum temperature, minimum temperature, precipitation, and etc and as it does not reflect temperature and humidity, uses the weather index association rule and correlation diagram to calculate the user based industry weather index. Present weather index assumes the industry weather index as the missing value and the user's industry weather index is forecasted through collaborative fil-

Fig. 4 Emerging risk forecast process using mining engine

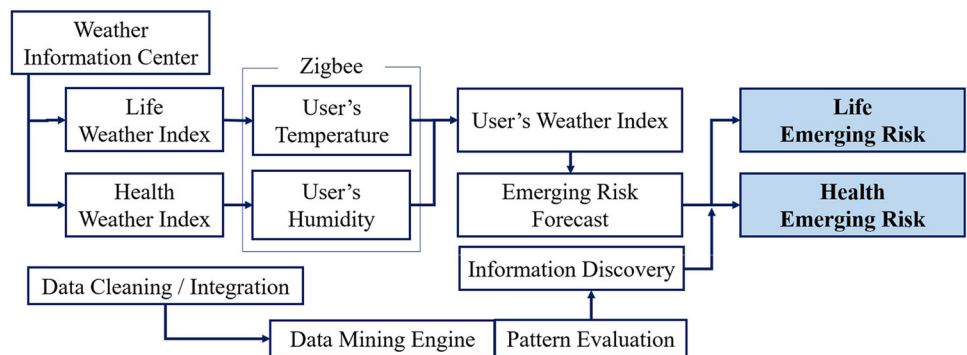


Fig. 5 Industry emerging risk forecast process using mining engine

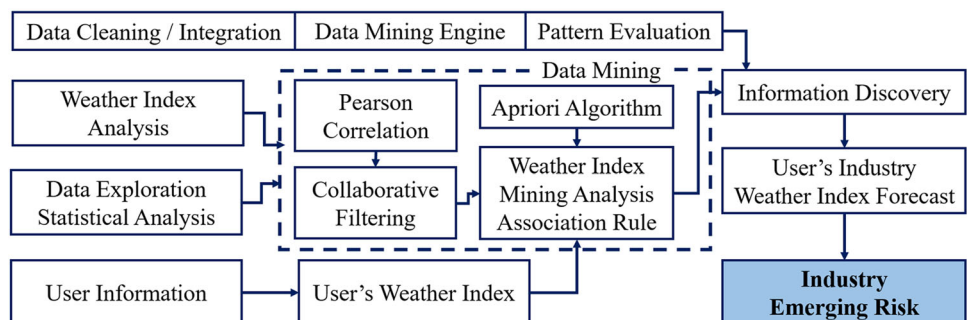


Table 6 Range of emerging risk forecast

Range	0–1.5	1.5–2.5	2.5–3.5	3.5–4.0
Emerging risk	Low	General	High	Very high

tering with correlation as its weighted value. The forecasted user’s industry weather index is scanned and the association rule is applied. The average value of the calculated user’s industry weather index is forecasted and the average value is forecasted as the industry emerging risk. Figure 5 shows the industry emerging risk forecast process using mining engine.

Emerging risk is forecasted dividing the average value of the user based weather index from the vector field from 0 to 4. The forecasted emerging risk is classified into low, general, high, and very high. Table 6 represents the range of emerging risk forecast.

4 System and implementation

4.1 Service environment

This study suggests the emerging risk forecast system using the associative index mining analysis. This is the user based weather index monitoring weather system that analyzes the weather index data and extracts the related relationship of the correlation between the index, reflect variables by the user situation and calculates the weather index. Weather data XML File is collected such as the weather index, village forecast, weather forecast etc is collected by OPEN API provided in the Meteorological Administration [14]. Collected weather data forms transaction and the association relationship and correlation between the index is analyzed using TANAGRA 1.4.50 Version [27] and the SPSS 24 version [29]. Weather index by the user situation is re-calculated by using the analyzed result, user temperature and user humidity. Calculated weather index is used to forecast the emerging risk and is

services to the user through the monitoring system. Figure 6 shows the service environment.

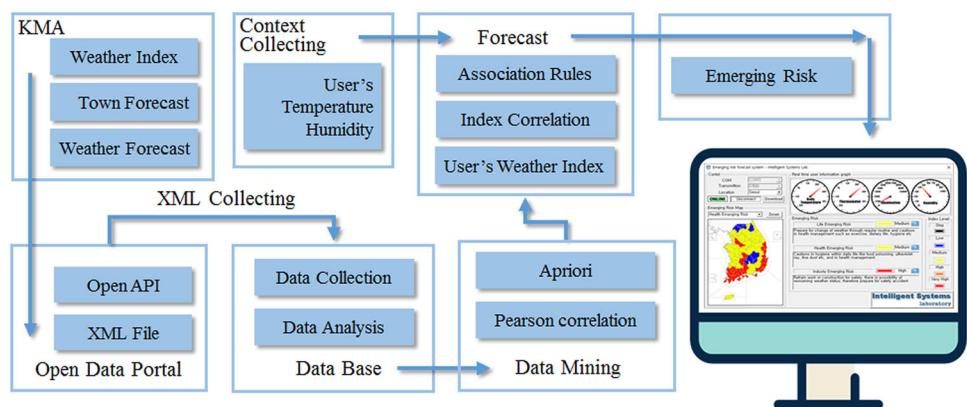
4.2 Development of emerging risk forecast system

Emerging risk forecast system by the situation of the user was developed by using the XML parser and Zigbee communication in the Window programming MFC based environment. Monitoring system is formed of communication information, weather information, weather index and emerging risk [2, 18]. Communication information part is the part collecting the user temperature and humidity and is formed of the user information table and graph collected real time and the connection status and the interface selecting the communication port, communication velocity, region and etc. Temperature, illumination intensity, humidity, and ultraviolet ray input by the present connected to the real-time sensor data field is shown in numbers. The real-time user information graph field is shown on the screen by visualizing the collected user information. Figure 7 shows the developing emerging risk forecast system.

Weather information part is shown by using the XML file collected through the RSS of the Meteorological Administration with the information of the region selected by the weather user. The collected XML file includes the weather, present temperature, maximum/minimum temperature, precipitation probability, precipitation forecast, snowfall amount forecast, wind speed etc. Longitude (x) and latitude (y) coordinate was input to the neighbor forecast signal address and collects the weather forecast of the XML file of the region [14].

Weather index part provides the life weather index, health weather index, industry weather index and the life weather index, health weather index is calculated using the information by the situation of the user. Industry weather index stopped its service in January 2016 and therefore it is directly calculated and provided from the program based on the weather information of the Meteorological Administration [16]. It provides along the weather index status and precautions according to the extent of the index. Emerging risk

Fig. 6 Service environment



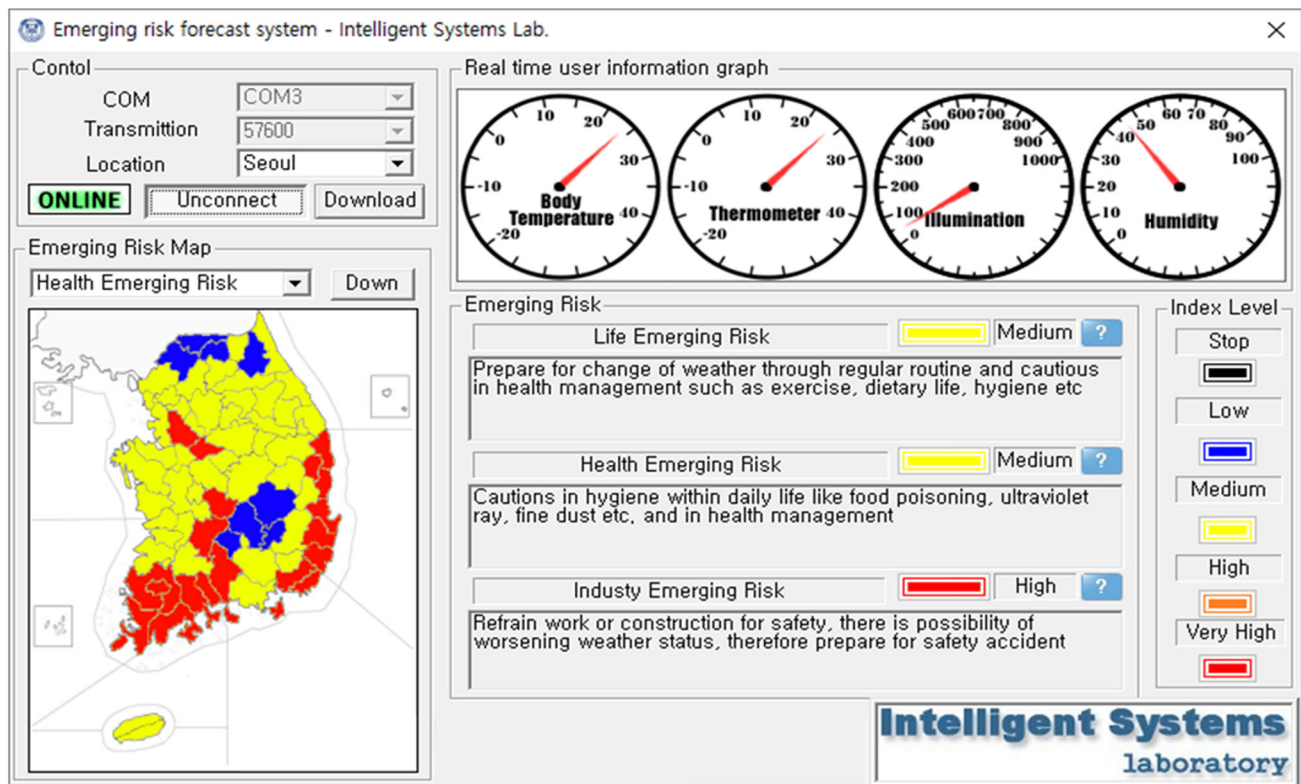


Fig. 7 Developing emerging risk forecast system

Table 7 Precautions of emerging risk per step

Emerging risk	Level	Color	Precautions
Life emerging risk	Low	Blue	No danger to be specially aware of but need to give continuous interest and cautious on health management
	General	Yellow	Cautions in hygiene within daily life like food poisoning, ultraviolet ray, fine dust etc. and in health management
	High	Orange	Cautious in health management in daily life as there is danger of food poisoning, heat wave, ultraviolet ray etc.
	Very High	Red	Cautious on skin exposure as there is high danger in food poisoning, heat wave, ultraviolet ray etc and abstain from going out and cautious on food
Health emerging risk	Low	Blue	Cautious on daily health management
	General	Yellow	Prepare for change of weather through regular routine and cautious in health management such as exercise, dietary life, hygiene etc.
	High	Orange	Maintain body temperature as there is possibility of occurring disease by external factors such as daily temperature difference, ultraviolet rays, flower dust etc. and cautious in health management
	Very High	Red	Maintain environment cleanly, cautious not to be exposed from dramatic weather change, abstain from outdoor activity
Industry emerging risk	Low	Blue	Good weather status and high industry efficiency, no special danger but prepare for safety accident
	General	Yellow	No special danger but there is possibility of low work efficiency, weather status not bad
	High	Orange	Refrain work or construction for safety, there is possibility of worsening weather status, therefore prepare for safety accident
	Very High	Red	Danger extent of the industry site is high due to many factors, therefore not adequate to work as there could be high probability of safety accident, prepare for damage in human life or asset

part is formed of life emerging risk, health emerging risk and industry emerging risk and provides precautions per step with the danger level. Table 7 shows the precautions per step of emerging risk.

Emerging risk forecast system is formed by comprehending the communication information, weather information, weather index and emerging risk. System operates in the order of sensor connection, communication port and velocity selection, region selection and download order. If the system and sensor is rightly connected it shows green on the left online and the real time measured temperature, humidity, illumination intensity and ultraviolet ray value on the real time sensor data field is expressed. Also present sensory temperature, temperature, illumination intensity is shown on the real time user information graph field dashboard. Weather information is collected and shown on the screen when clicking on the download button. Overall weather information and weekly weather forecast is shown through pop-up screen when clicking the whole screen and weather forecast. The message box on the below right shows the communication status.

5 Conclusions

As information technology has developed the field of information widened and communication became possible without limit to time and place. Government disclosed to the people the public data such as existing observation data, statistics and etc to the people and is implementing various studies using this. Extent of influence by the weather to the life, health and industry is calculated by analyzing the weather data and weather index per region is calculated using the data collected real time from the regional weather observation point. Weather index service is provided through the internet and mobile app and the present weather index service is regional based service facing limits to the service by the personal situation. This study suggests emerging risk forecast system using associative index mining analysis. The suggesting method is the monitoring system focused on the user by analyzing the associative index using data mining and forecasting emerging risk by reflecting the user situation information. Emerging risk is classified into life emerging risk, health emerging risk, industry emerging risk by the forecast method. Emerging risk is divided into four steps of low, general, high and danger and as it gets low it means there is smaller probability of occurring damage to human life or asset. Collected data for emerging risk forecast were the weather index XML file of the Meteorological Administration and the user situation information. The collected weather index was to analyze the association and related relationship between the index using Apriori algorithm of data mining and *Pearson* correlation coefficient. User based

weather index was calculated using the analysis result and user situation information and this was then used to forecast the emerging risk. Monitoring system was developed including precautions followed by the forecasted emerging risk and weather index, weather information, user situation information etc. Emerging risk forecast system using the developed associative index mining analysis is the personalized monitoring system which complemented the disadvantage of the weather index that was limited to the existing region.

Acknowledgements This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1A09917313).

References

- Zun, I., Chung, K.Y.: Life weather index monitoring system using wearable based smart cap. *J. Korea Contents Assoc.* **9**(12), 477–484 (2009)
- Jung, H., Chung, K.: Life style improvement mobile service for high risk chronic disease based on PHR platform. *Clust. Comput.* **19**(2), 967–977 (2016)
- Jung, H., Chung, K.Y.: Ontology-driven slope modeling for disaster management service. *Clust. Comput.* **18**(2), 677–692 (2015)
- Chun, H.W.: Disaster prevention information technology. *Electron. Telecommun. Trends* **28**(2), 145–154 (2013)
- Kim, W.I., Park, W.G.: Present state and prospect of wireless networks for public protection and disaster relief. *Electron. Telecommun. Trends* **26**(3), 50–60 (2011)
- Disaster Risk Index: United Nations Development Programme. <https://www.undp.org/>
- Climate Change Vulnerability Index: Verisk Maplecroft. <https://maplecroft.com/>
- Jung, H., Chung, K.: Knowledge-based dietary nutrition recommendation for obese management. *Inf. Technol. Manag.* **17**(1), 29–42 (2016)
- Chung, K., Park, R.C.: P2P cloud network services for IoT based disaster situations information. *Peer Peer Netw. Appl.* **9**(3), 566–577 (2016)
- Yoo, H., Chung, K.: PHR based diabetes index service model using life behavior analysis. *Wirel. Pers. Commun.* (2016). doi:[10.1007/s11277-016-3715-9](https://doi.org/10.1007/s11277-016-3715-9)
- Chung, K., Oh, S.Y.: Voice activity detection using improvement unvoiced feature normalization process in noisy environment. *Wirel. Pers. Commun.* **89**(3), 747–759 (2016)
- Jung, H., Chung, K.: PHR based life health index mobile service using decision support model. *Wirel. Pers. Commun.* **86**(1), 315–332 (2016)
- Kim, J.C.: Mining based Potential risk monitoring system using weather data analysis. Sangji University, Master Paper (2016)
- Korea Meteorological Administration. <http://web.kma.go.kr/eng/>
- Jung, H., Chung, K.: P2P context awareness based sensibility design recommendation using color and bio-signal analysis. *Peer Peer Netw. Appl.* **9**(3), 546–557 (2016)
- Open Data Portal. <http://www.data.go.kr/>
- Jo, S.M., Chung, K.Y.: Design of access control system for telemedicine secure XML documents. *Multimed. Tools Appl.* **74**(7), 2257–2271 (2015)

18. Chung, K., Park, R.C.: PHR open platform based smart health service using distributed object group framework. *Clust. Comput.* **19**(1), 505–517 (2016)
19. Ahn, Y.H., Han, B., Ryu, Y., Bae, J.: Research of developing the National Safety & Security Index. National Disaster Management Research Institute (2014)
20. National Disaster Management Research Institute. <http://www.ndmi.go.kr/>
21. Chung, K.: Development of Ontology based Intelligent Slide Modeling and Urban Climate Disaster Index, Final R&D Report, 11-1613000-001562-01. Infrastructure and Transportation Technology Promotion Research Program, Korea Ministry of Land, Infrastructure and Transport (2016)
22. Wei, W., Guanyu, H., Junli, W.: Research on Zigbee wireless communication technology. In: Proceedings of International Conference on Electrical and Control Engineering, pp. 1245–1249 (2011)
23. Kim, J.C., Jung, H., Chung, K.: Mining based urban climate disasters index service according to potential risk. *Wirel. Pers. Commun.* **89**(3), 1009–1025 (2016)
24. Agrawal, R., Srikant, R.: Mining Sequential Patterns. In: Proceedings of the International Conference on Data Engineering, Data Engineering, pp. 3–14 (1995)
25. Jung, H., Chung, K.Y.: Mining based associative image filtering using harmonic mean. *Clust. Comput.* **17**(3), 767–774 (2014)
26. Hong, S.K., Kim, S.P., Lim, H.S., Moon, Y.S.: Secure multi-party computation of Pearson correlation coefficients. *J. KIISE* **41**(10), 799–809 (2014)
27. Tanagra. <http://data-mining-tutorials.blogspot.kr/>
28. Lee, B.W.: Digital convergence teaching strategy system using spearman correlation coefficients. *J. Internet Comput. Serv.* **11**(6), 111–122 (2010)
29. SPSS 24. <http://www.ibm.com/>
30. Kim, S.H., Chung, K.: Emergency situation monitoring service using context motion tracking of chronic disease patients. *Clust. Comput.* **18**(2), 747–759 (2015)



Joo-Chang Kim has received B.S. and M.S. degrees from the School of Computer Information Engineering, Sangji University, Korea in 2014 and 2016, respectively. He has worked for Network Management Department, Network O&S Co., Ltd. He is currently in the doctorate course of the Department of Computer Information Engineering, Sangji University, Korea. He has been a researcher at Intelligent System Lab., Sangji University. His research interests include Data Mining, Healthcare, Medical Decision System, HCI, and Recommendation.



Kyungyong Chung has received B.S., M.S., and Ph.D. degrees in 2000, 2002, and 2005, respectively, all from the Department of Computer Information Engineering, Inha University, Korea. He has worked for Software Technology Leading Department, Korea IT Industry Promotion Agency (KIPA). He is currently a professor in the School of Computer Information Engineering, Sangji University, Korea. His research interests include Medical Data Mining, Healthcare, Knowledge System, HCI, and Recommendation.