

Available online at www.sciencedirect.com**ScienceDirect**journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Multidimensional analysis model for highly pathogenic avian influenza using data cube and data mining techniques



Zhenshun Xu ^{a,b}, Jonguk Lee ^a, Daihee Park ^{a,*}, Yongwha Chung ^a

^a Department of Computer Convergence Software, College of Science and Technology, Korea University, Sejong Campus, 2511 Sejong-ro, Sejong 30019, Republic of Korea

^b School of Computer Science and Engineering, Beifang University of Nationalities, No.204 Wenchang, North-Street, Xixia District, Yinchuan, Ningxia 750021, China

ARTICLE INFO

Article history:

Received 25 May 2016

Received in revised form

25 February 2017

Accepted 14 March 2017

Published online 31 March 2017

Keywords:

Highly pathogenic avian influenza

Data cube model

OLAP analysis

Association rule mining

Sequential pattern mining

The highly pathogenic avian influenza (HPAI) viral disease can spread rapidly, resulting in high mortality rates and severe economic damage. To minimize the damage incurred from such diseases, it is necessary to develop technology for collecting and analysing livestock disease data. In this paper, we propose a data cube model with data mining techniques for the analysis of HPAI using livestock disease data accumulated over an extended period. Based on the construction of the data cube model, a multidimensional HPAI analysis is performed using online analytical processing (OLAP) operations to assess the temporal and spatial perspectives of the spread of the disease with varying levels of abstraction. Furthermore, the proposed analysis model provides useful information that generates site connectedness and potential sequential dissemination routes of HPAI outbreaks by applying association rule mining and sequential pattern mining, respectively. We confirm the feasibility and applicability of the proposed HPAI analysis model by implementing and applying an analysis system to HPAI outbreaks in South Korea.

© 2017 IAgE. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Over the past decades, highly infectious livestock diseases have increased threats to human life, causing considerable economic damage and environmental problems (OIE, 2016). In particular, several outbreaks of highly pathogenic avian influenza (HPAI) have been reported in South Korea, since the first detection of the HPAI virus at the end of 2003 (Kim, 2015; Kim et al., 2015; Shin et al., 2015). Even though no cases of livestock-to-human transmission of HPAI have been reported

in South Korea, the occurrence of this disease in other countries has resulted in high fatality rates in humans (Tuncer & Le, 2014; WHO, 2016). To minimize the damage resulting from such diseases, it is necessary to develop technology for analysing livestock disease occurrence data.

In general, migratory birds are major candidates for the long-distance dispersal of HPAI (Kwon et al., 2016; Verhagen, Herfst, & Fouchier, 2015). When HPAI initially occurs, it can spread to short distances by the methods of direct and indirect transmission (Seo et al., 2014). According to the latest manual

* Corresponding author. Fax: +82 44 860 1584.

E-mail address: dhpark@korea.ac.kr (D. Park).

<http://dx.doi.org/10.1016/j.biosystemseng.2017.03.004>

1537-5110/© 2017 IAgE. Published by Elsevier Ltd. All rights reserved.

Table 1 – Summary of HPAI outbreak status.

Outbreak year (starting season)	Virus type	No. of diagnosed farms (by year)	Location of first detected site (state.city)	No. of outbreak sites (city town)
2003/2004 (winter)	H5N1	14/5	Chungbuk.Eumseong	10 14
2006/2007 (winter)	H5N1	4/3	Jeonbuk.Iksan	5 6
2008 (spring)	H5N1	33	Jeonbuk.Gimje	19 26
2010/2011 (winter)	H5N1	2/51	Jeonbuk.Iksan	25 39
2014/2015 (winter)	H5N8	250/120	Jeonbuk.Gochang	57 127

Table 2 – Statistical summary of poultry farms and animals in South Korea in 2015.^a

Breeding States or cities	Broiler chicken		Layer chicken		Duck		Total	
	No. farms	No. animals	No. farms	No. animals	No. farms	No. animals	No. farms	No. animals
Seoul city	15	20,947	36	739,948	4	2424	55	763,319
Busan city	34	14,997	105	433,004	27	12,170	166	460,171
Daegu city	90	97,578	155	604,461	7	57	252	702,096
Yinchen city	165	512,706	389	448,223	33	1335	587	962,264
Guangju city	48	364,888	58	73,688	18	96,480	124	535,056
Daejeon city	88	186,741	77	1,002,511	5	70	170	1,189,322
Ulsan city	117	7759	292	342,806	41	792	450	351,357
Sejong city	77	731,124	129	990,927	9	94	215	1,722,145
Gyeonggi state	1115	13,661,435	2133	10,753,361	181	490,253	3429	24,905,049
Gangwon state	1704	4,457,766	2834	2,027,658	166	38,055	4704	6,523,479
Chungbuk state	745	6,289,535	1105	2,005,064	126	586,445	1976	8,881,044
Chungnam state	1161	21,546,357	1584	5,960,939	214	329,446	2959	27,836,742
Jeonbuk state	1016	19,512,154	1076	4,403,169	187	3,445,601	2279	27,360,924
Jeonnam state	1781	13,351,440	2016	2,932,017	328	2,774,418	4125	19,057,875
Gyeongbuk state	1467	11,525,418	2148	9,673,861	142	77,035	3757	21,276,314
Gyeongnam state	1274	4,858,531	2326	5,712,178	202	450,807	3802	11,021,516
Jeju state	42	1,690,872	41	416,590	8	1619	91	2,109,081
Total	10,939	98,830,248	16,504	48,520,405	1698	8,307,101	29,141	155,657,754

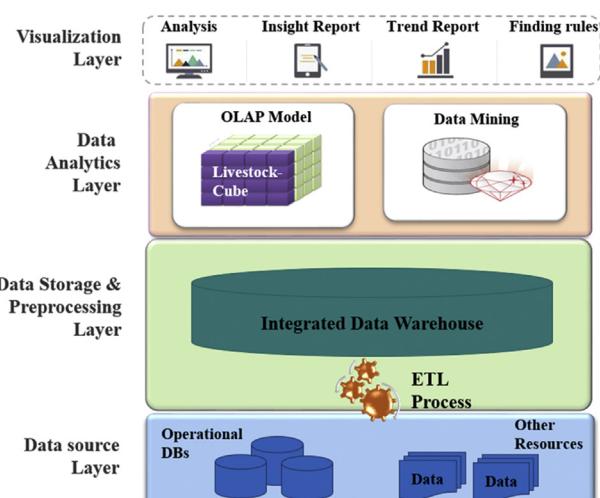
^a Source: http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1AG15501&conn_path=I3.

of control measures in South Korea, HPAI outbreaks in poultry have been controlled by a variety of possible measures such as movement controls, stamping-out, preemptive slaughter of poultry on adjacent sites and cleaning and disinfection of infected sites (MAFRA, 2016). However, in spite of various strategic efforts to control the spread of HPAI, these emerging viruses continue to survive and evolve (OIE, 2016). Thus, early preventive measures against epidemics are the most important methods to reduce the damage of an outbreak.

To date, an extensive variety of studies have reported on HPAI outbreaks. Seo et al. (2014) used computational fluid dynamics to estimate the dispersion of the virus attached to aerosols produced by livestock. They used a geographical information system to model a specific three-dimensional topography that included farm locations, road networks, and related facilities. Lee et al. (2014) constructed a direct HPAI spread network based on the relationships between farms by using poultry-related business data and an indirect HPAI spread network using the aerial spread from each farm during the HPAI outbreak in 2008. Tuncer and Le (2014) studied the effects of air travel on the spread of avian influenza from Asian and Australian cities to the United States. A two-city mathematical model involving a pandemic strain was used to derive the basic reproduction number, which determined whether the disease would spread and persist or become extinct. Real air travel data was used to model the spread of

the disease by individuals who were susceptible to or were infected with pandemic avian influenza.

Unlike the perspectives in the current research literature, in this study, we propose a data cube model with data mining techniques for the multidimensional analysis of HPAI using livestock disease data provided by the Ministry of Agriculture in South Korea (KAHIS, 2016). Based on the construction of the

**Fig. 1 – Overall structure of the HPAI analysis system.**

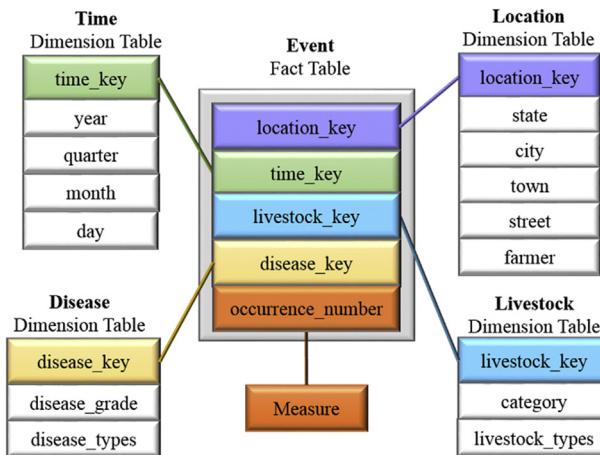


Fig. 2 – Star schema of livestock disease analysis model.

data cube model, multidimensional HPAI analysis is then performed using online analytical processing (OLAP) operations to assess the temporal and spatial perspectives with varying levels of abstraction. Furthermore, the proposed analysis model provides detailed information of long-term HPAI data by using data mining techniques such as association rule mining and sequential pattern mining. We describe a sample process based on a specific scenario eliciting information that generates site connectedness and potential sequential dissemination routes of HPAI outbreaks by applying association rule mining and sequential pattern mining, respectively. To the best of our knowledge, this is the first report of an OLAP analysis for accumulated long-term livestock disease occurrence data and data mining analysis for the in-depth analysis of disease outbreaks. We test the feasibility and applicability of the proposed HPAI analysis

model by implementing an HPAI analysis system and applying it to the analysis of HPAI outbreaks in South Korea.

2. Materials and methods

2.1. Data sources

In our experiments, we used the livestock disease occurrence data provided by the Ministry of Agriculture in South Korea over 13 years from 2003 to June 2015. Since the first detection of the HPAI virus at the end of 2003, several occurrences of HPAI have been reported in South Korea (KAHIS, 2016). Table 1 presents a summary of HPAI occurrences across the country, including the outbreak year, virus type, number of diagnosed farms, location of the first detected outbreak site, and total number of outbreak sites. Recent studies indicated there were five distinct occurrences of HPAI (2003–2004, 2006–2007, 2008, 2010–2011 and 2014–2015) in South Korea during the last decade (Kim, 2015; Kim et al., 2015; Shin et al., 2015; Yoon et al., 2015). In addition, Table 2 provides a statistical summary of poultry farms in South Korea in 2015 based on a survey conducted every five years (the latest investigation was conducted in 2015). Note that the major type of facilities where poultry is raised is a commercial factory where poultry have limited access to wild birds.

2.2. Proposed HPAI occurrence analysis system

In this section, we introduce an HPAI data cube analysis model that can analyse long-term HPAI occurrence data using various OLAP operations at multiple levels of abstraction. Figure 1 illustrates the overall structure of the proposed analysis system. As shown in this figure, the system consists

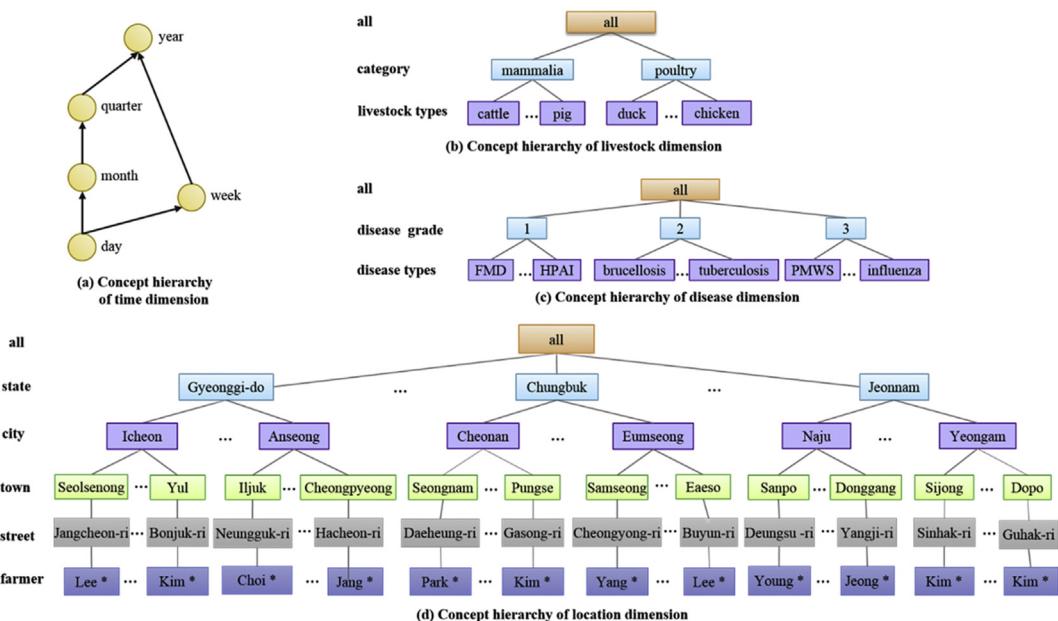


Fig. 3 – Concept hierarchy of (a) time dimension; (b) livestock dimension; (c) disease dimension; (d) location dimension.

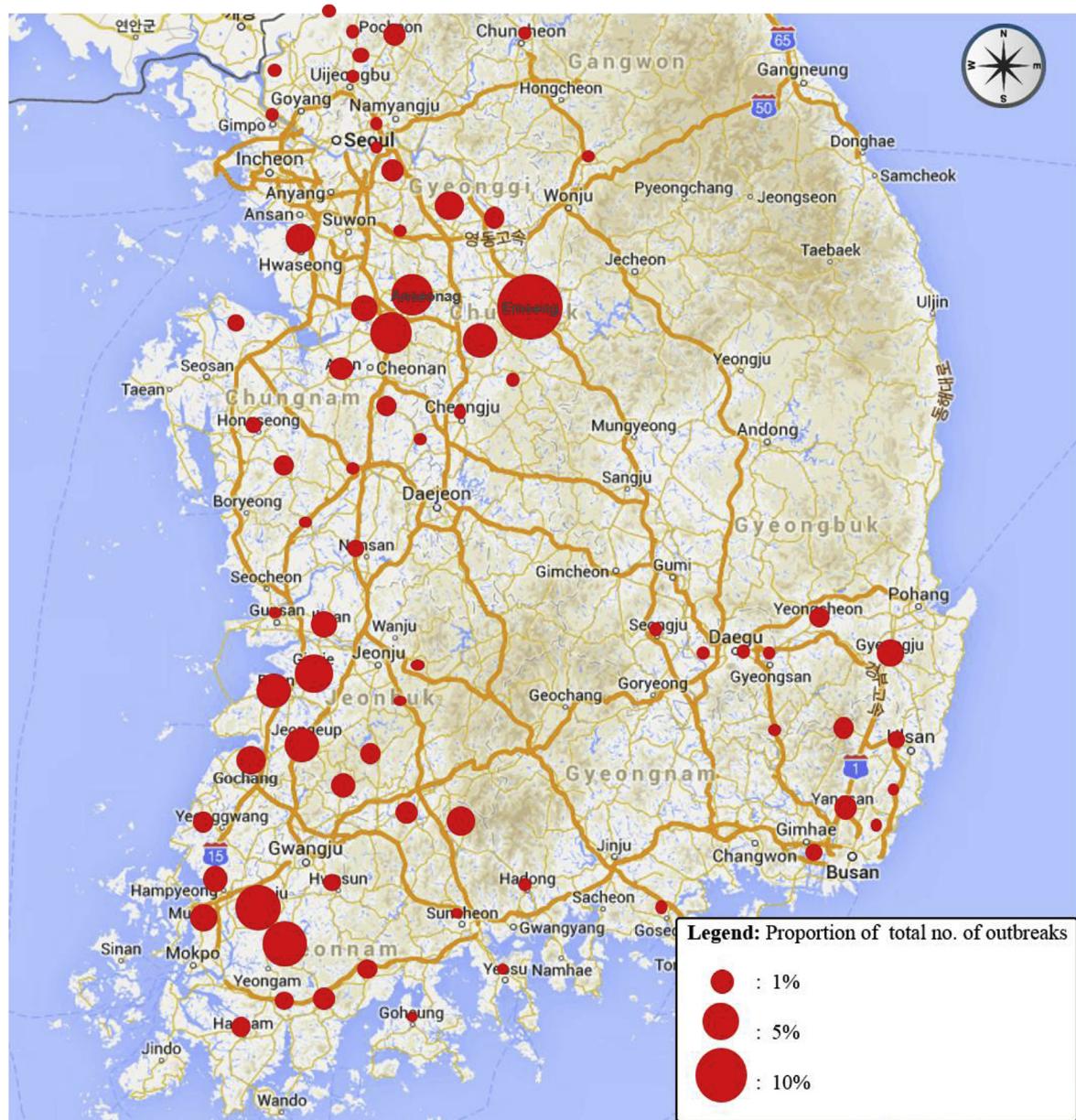


Fig. 4 – HPAI outbreak rate in each city across the country through 2003–2015.

of four parts: source data repositories, a data warehouse, data analytics with OLAP and data mining techniques, and visualization. The various forms of data repositories in the first layer are delivered to the next layer of storage. All data related to livestock disease occurrence data, route of migratory birds, statistical data, and weather data might be candidate data repositories.

To ensure that only data in the standard format is stored in the data warehouse, a process called extract, transform, and load (ETL) is performed to preprocess the data collected from the different data sources. In the third layer, we perform multidimensional analysis using OLAP and data mining techniques. The last visualization layer interface provides valuable information for the development of a disease prevention and control strategy for the country.

Each module will be described in detail.

2.3. Multidimensional data cube

A data warehouse requires a concise, subject-oriented schema that facilitates online data analysis. The most popular data model for a data warehouse is the multidimensional model. This model represents data in the form of a data cube. A data cube allows data to be modelled and viewed in multiple dimensions. It is defined by dimensions and facts. In general, dimensions are the perspectives or entities with respect to which a user wishes to maintain records. The fact table contains the names of the facts, or measures, and keys to each of the related dimension tables (Han, Kamber, & Pei, 2012, chap. 4 and 6). In this paper, we used a star schema for the multidimensional analysis model associated with four dimensions (time, location, disease, and livestock) and “occurrence number” as the measurement in

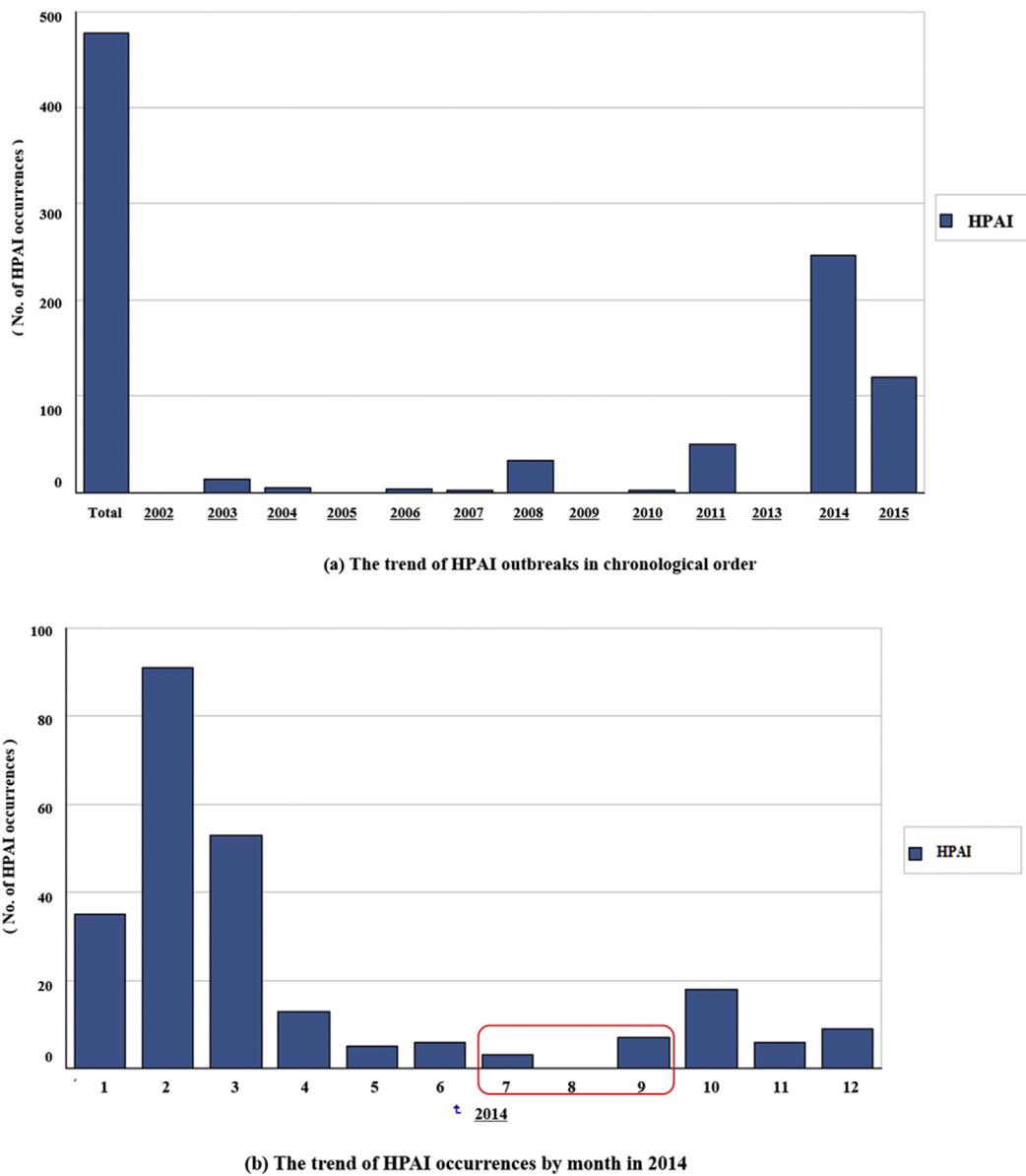


Fig. 5 – (a) Trend of HPAI outbreaks in chronological order; (b) The trend of HPAI occurrences by month in 2014.

the fact table. Figure 2 illustrates the star schema for the proposed model.

2.4. Concept hierarchy and OLAP operations

A concept hierarchy defines a sequence of mappings from a set of low-level concepts to high-level and more generalized concepts. It allows data to be managed at varying levels of abstraction. The attributes of the time dimension are organized in a partial order, forming a lattice. The partial order for the time dimension based on attributes such as day, week, month, quarter, and year is “day < {month < quarter; week} < year” (see Fig. 3(a)). The concept hierarchy for the location dimension is based on the structure of a hierarchical tree, from the lowest level “farmer” to the highest level “state” (see Fig. 3(d)). Similarly, the concept hierarchies for livestock and disease are presented in Fig. 3(b) and (c), respectively.

A number of OLAP data cube operations exist in the multidimensional model, allowing interactive querying and analysis of the available data. OLAP provides a user-friendly environment for interactive data analysis, which makes it possible to construct a flexible and extensible analysis system. Several OLAP operations such as roll-up, drill-down, slice, and dice are used to retrieve useful information from the HPAI data warehouse. The roll-up operation performs aggregation on a data cube, either by climbing a concept hierarchy for a dimension or by dimension reduction. The drill-down operation is the reverse of the roll-up operation; it navigates from less detailed data to more detailed data. The slice operation selects one dimension of a given cube resulting in a subcube. The dice operation defines a subcube by performing a selection on two or more dimensions (Han et al., 2012, chap. 4 and 6). For the deployment of the OLAP module, Oracle analytic workspace manager (AWM) was employed to build the data

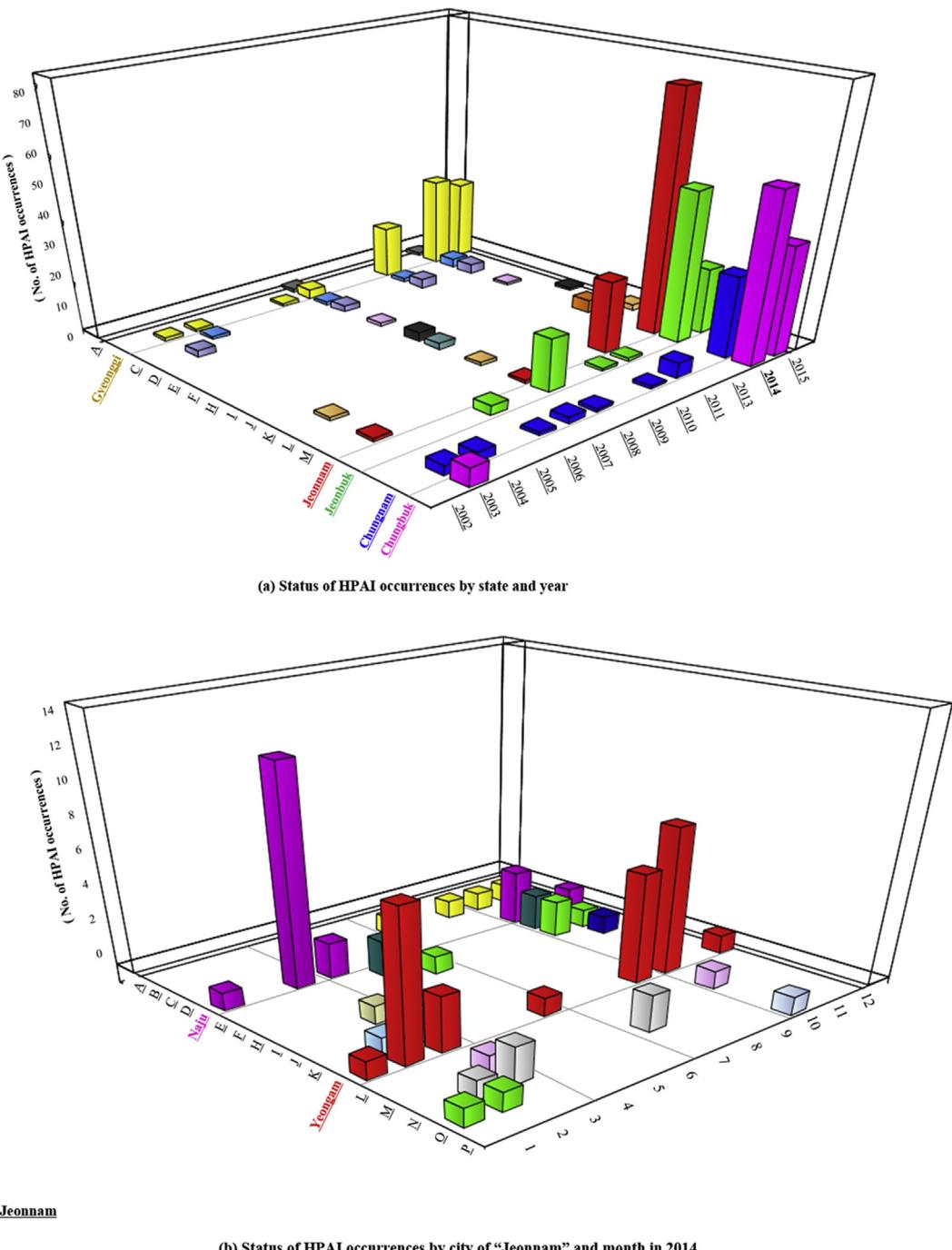


Fig. 6 – Temporal and spatial analysis of HPAI outbreaks. (a) Status of HPAI occurrences by state and year; (b) Status of HPAI occurrences by city of “Jeonnam” and month in 2014.

cube by connecting it to the data warehouse ([Oracle AWM](#), 2016). After being loaded into the data warehouse, the data that were crucial for analysis were selected from the data warehouse and imported into the OLAP. The data cube was built in the OLAP server for analysis and reporting. As the dimension tables contain hierarchical data, data are presented in different hierarchies so as to facilitate a multidimensional view of the OLAP. As all calculations are pre-computed in the OLAP server, user can analyse the data or

create reports in real-time, depending on what is needed, by choosing the desired dimensions.

2.5. Data mining techniques

In this section, we briefly describe two data mining techniques used to elicit useful information for an in-depth understanding of HPAI diseases: association rule mining and sequential pattern mining.

Table 3 – Example of obtained association rules.

No.	Association rule (state.city)
1	Jeonbuk.Gochang → Jeonbuk.Iksan
2	Jeonbuk.Gochang → Jeonnam.Naju Jeonnam.Yeongam
3	Jeonbuk.Gochang → Chungnam.Asan Chungnam.Chenan
4	Jeonbuk.Gochang → Gyeonggi.hwaseong Gyeonggi.Ichen Gyeonggi.Pyeongtak Gyeonggi.Anseong
5	Jeonbuk.Gochang → Gyeongbuk.Yangsan

Association rule mining aims at uncovering interesting and useful patterns in enormous amounts of data. It consists of first determining frequent itemsets (a set of items, such as A and B, satisfying a minimum support threshold, or percentage

of the task-relevant tuples) from which strong association rules in the form of $A \Rightarrow B$ are generated. These rules also satisfy a minimum confidence threshold (a prespecified probability of satisfying B under the condition that A is satisfied). Associations can be further analysed to uncover correlation rules, which convey statistical correlations between itemsets A and B (Han et al., 2012, chap. 4 and 6). For the association rule mining in this study, we used “arules” package in R tools (R arules, 2015). It provides the infrastructure for representing, manipulating, and analysing transaction data and patterns, as well as interfaces to C implementations of the association mining algorithm Apriori (R arules, 2015; Agrawal, Imielinski, & Swami, 1993).

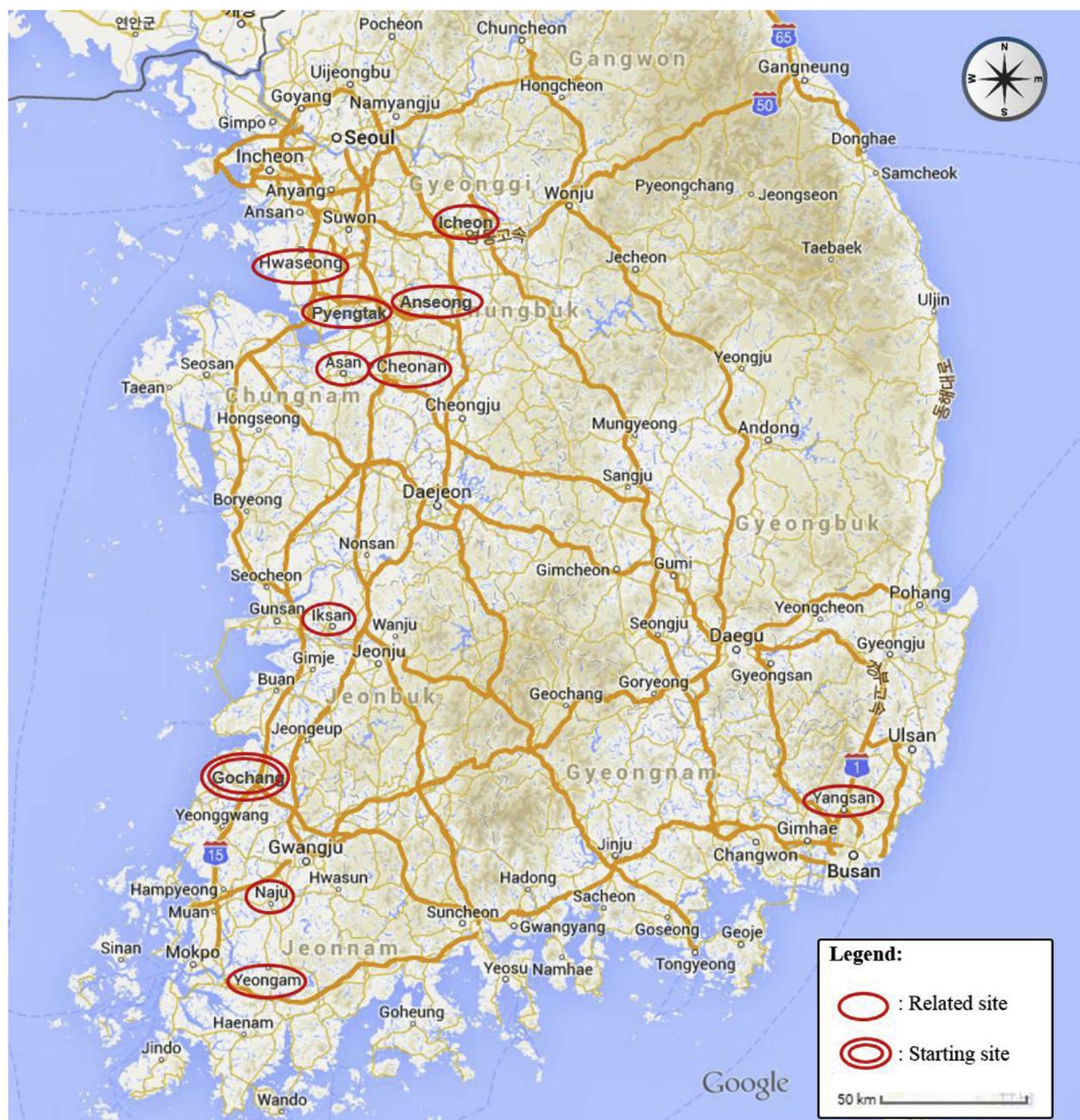
**Fig. 7 – Site connectedness of HPAI outbreaks with “Gochang”.**

Table 4 – Example of obtained sequential pattern rules with city level.

No.	Sequential route of HPAI outbreaks (state.city)
1	Jeonbuk.Gochang → Jeonbuk.Buan
2	Jeonbuk.Gochang → Jeonnam.Haenam
3	Jeonbuk.Gochang → Jeonnam.Naju → Jeonnam.Yeongam
4	Jeonbuk.Buan → Chungnam.Buyeo Jeonbuk.Jeongeup
5	Chungnam.Buyeo → Chungnam.Chenan → Gyeonggi-do.Hwaseong
6	Chungnam.Chenan → Chungbuk.Jinchen → Chungbuk.Emseong
7	Chungnam.Chenan → Gyeonggi-do.Anseong → Gyeonggi-do.Pyeongtak
8	Jeonbuk.Jeongeup → Jeonbuk.Imsil → Gyeongnam.Miryang
9	Jeonbuk.Jeongeup → Jeonbuk.Gimje → Chungnam.Nonshan
10	Jeonnam.Yeongam → Jeonnam.HamPyeong → Jeonnam.Yeonggwang

Sequential pattern mining is the discovery of frequently occurring ordered events or subsequences as patterns. Given a set of sequences, where each sequence consists of a list of events (or elements) and each event consists of a set of items, and a user-specified minimum support threshold, sequential pattern mining identifies all the frequent subsequences; that is, the subsequences whose frequency of occurrence in the set of sequences is no less than a minimum support threshold (Agrawal & Srikant, 1995). For sequential pattern mining in this study, we used “arulesSequences” package in R tools (R arulesSequences, 2015). It is an add-on for arules to handle and mine frequent sequences and provides interfaces to the C++ implementation of cSPADE (R arulesSequences, 2015; Zaki, 2001).

3. Results and discussions

In this section, we describe a multidimensional HPAI outbreak analysis system. Further, we test the feasibility and applicability of the proposed system by describing the experimental results that were applied to the HPAI outbreak data provided by the Ministry of Agriculture in South Korea using Oracle AWM (Oracle AWM, 2016) and R tools (R arules, 2015; R arulesSequences, 2015). Figure 4 illustrates the HPAI

outbreaks in South Korea over 13 years, from 2003 to June 2015. Here, the size of the circle refers to the proportion of the total number of outbreaks in each city of South Korea.

3.1. HPAI analysis results using OLAP

This section introduces multidimensional analysis results that can be obtained by choosing a dimension within the viewpoint that a system manager is attempting to analyse, and also by adjusting the level of abstraction using the proposed model described in the previous section. First, we demonstrate that the system can be used to perform a descriptive analysis of the HPAI occurrences in space and time in order to identify current trends and location patterns. The following formula is a query statement with a symbolic meaning that is used to analyse the changes in HPAI outbreak trends according to the time slots:

Dice for (disease = “HPAI”) and (time = “each year”) and (location = “all”) and (livestock = “poultry”) and (measure = “occurrence_number”) (1)

Figure 5(a) presents the graph that resulted from formula (1). The graph indicates the trend of HPAI outbreaks in chronological order. In particular, Fig. 5(a) indicated that the recent HPAI outbreak (2014–2015) was more serious than previous

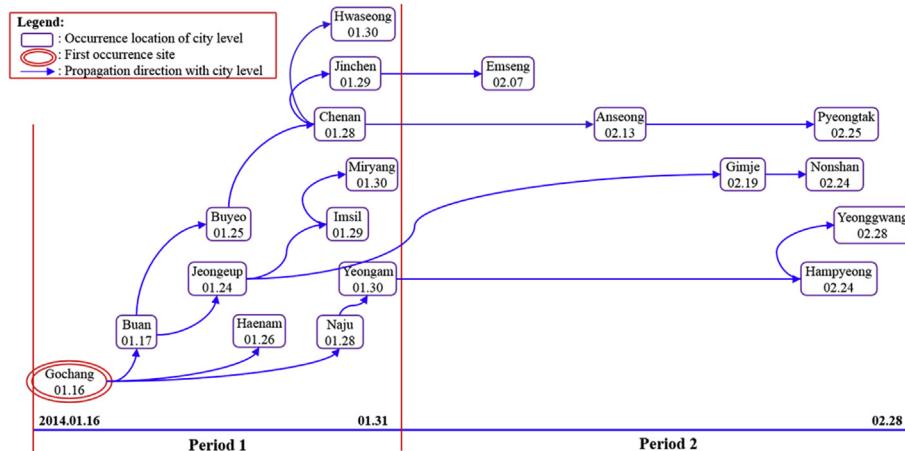


Fig. 8 – Time semantic graph of HPAI propagation routes from “Gochang” by city level.



Fig. 9 – Sequential propagation routes of HPAI outbreaks from “Gochang” in GIS.

occurrences, causing more concern. Thus, we then conducted a drill-down operation for the time dimension, as indicated in formula (2), in order to understand the results of the detailed analysis. Figure 5(b) presents the results of the operation and shows that it is consistent with the fact that most cases occur in winter and spring with no cases reported in August. Note that HPAI re-occurred in South Korea during fall 2014–winter 2015 following the initial introduction in winter 2014. Although the growing HPAI outbreak in fall 2014 coincided with the fall migration of migratory waterfowl, further phylogenetic analyses suggested that the HPAI virus detected

on poultry farms in late 2014 was not related to the viruses carried by wild waterfowl, but have instead evolved from the viruses circulating on poultry farms or among resident wild birds in South Korea since early 2014 (Kwon et al., 2016; Kim, 2015).

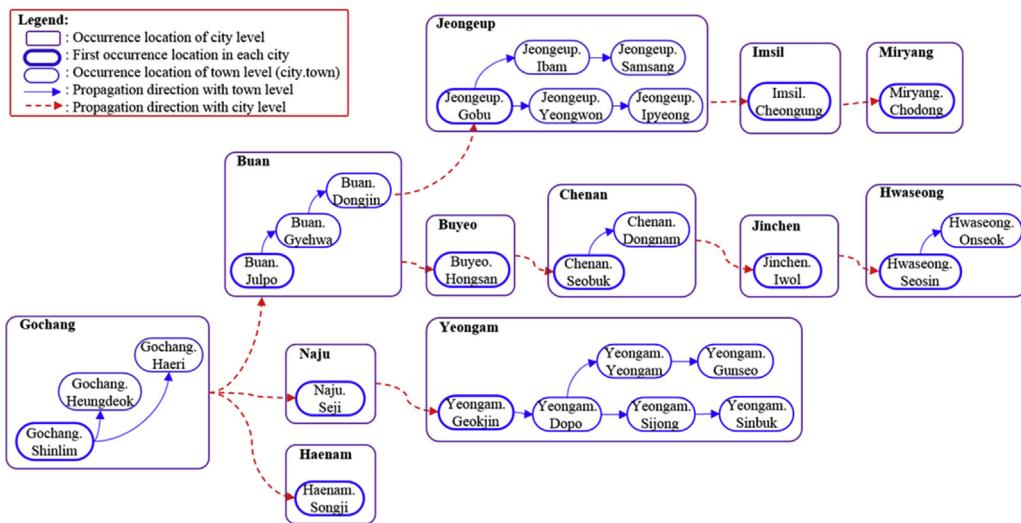
Drill-down on time (year to month)

Dice for(disease = “HPAI”) and (time = “each month in 2014”) and (livestock = “poultry”) and (location = “all”) and (measure = “occurrence_number”)

(2)

Table 5 – Example of obtained sequential pattern rules with town level.

No.	Sequential route of HPAI outbreaks (state.city.town)
1	Jeonbuk.Gochang.Shinlim → Jeonbuk.Gochang.Heungdeok Jeonbuk.Gochang.Haeri
2	Jeonbuk.Gochang.Shinlim → Jeonbuk.Buan.Julpo
3	Jeonbuk.Buan.Julpo → Chungbuk.Buan.Gyehwa → Chungbuk.Buan.Dongjin
4	Jeonbuk.Buan.Julpo → Jeonbuk.Buyeo.Hongsan
5	Jeonbuk.Buan.Julpo → Jeonbuk.Jeongeup.Gobu
6	Jeonbuk.Jeongeup.Gobu → Jeonbuk.Jeongeup.Yeongwon → Jeonbuk.Jeongeup.Ipyeong
7	Jeonbuk.Jeongeup.Gobu → Jeonbuk.Jeongeup.Ibam → Jeonbuk.Jeongeup.Samsang
8	Jeonbuk.Jeongeup.Gobu → Jeonbuk.Imsil.Cheongung → Gyeongnam.Miryang.Chodong
9	Jeonbuk.Gochang.Haeri → Jeonbuk.Naju.Seji
10	Jeonbuk.Buyeo.Hongsan → Jeonbuk.Chenan.Seobuk
11	Jeonbuk.Chenan.Seobuk → Jeonbuk.Chenan.Dongnam

**Fig. 10 – Location semantic graph of HPAI propagation routes from “Gochang.Shinlim” by town level.**

In terms of temporal and spatial analyses, a system manager can conduct a drill-down operation, as indicated in formula (3), to the state level of the location dimension and year level of the time dimension in order to confirm in detail which states of South Korea suffered from HPAI in the previous decades. The result of HPAI occurrences in each state indicates the number of times HPAI outbreaks occurred in each state and each year, as presented in Fig. 6(a). We can observe that “Jeonnam,” “Jeonbuk,” and “Chungbuk” experienced occurrences of HPAI several times in 2014. To perform a detailed temporal and spatial analysis, we drilled down from state level to the town level for the location dimension and from the year level to the month level in 2014 for the time dimension, as illustrated in formula (4). In Fig. 6(b), we can observe that HPAI occurred many times in “Yeongam” and “Naju.” In particular, the number of occurrences increased in “Yeongam” in September and October 2014.

Drill-down on location (all to state)
 Dice for
 (disease = “HPAI”) and (time = “each year”) and
 (livestock = “poultry”) and (location = “each state”)
 and (measure = “occurrence number”)

Drill-down on time (year to month)
 Drill-down on location (cities in “Jeonnam” state)
 Dice for (disease = “HPAI”) and (time = “each month in 2014”) and
 (livestock = “poultry”) and
 (location = “each city in “Jeonnam””) and
 (measure = “occurrence_number”)

(4)

3.2. HPAI analysis results using association rule mining

In this section, based on the temporal and spatial analysis results from the livestock data cube analysis model, we describe a process for the retrieval of useful information. This process works by the application of the association rule mining for the establishment of site connectedness of regions where HPAI occurs. As an example, in this paper, we focused on the first outbreak occurring in “Gochang” in order to explore long-distance spread of HPAI based on HPAI outbreaks data over 13 years from 2003 to June 2015. Table 3 presents a sample of the obtained association rules (support ≥ 0.4 and confidence = 1) without considering the sites within and adjacent to Gochang

(to focus on long-distance spread). Moreover, we depicted the HPAI co-occurrence sites with “Gochang” in a geographic information system (GIS) representation of South Korea (see Fig. 7). In Fig. 7, we can easily see that the next potential sites of an HPAI outbreak could be in “Iksan,” “Yeongam,” “Asan”, “Yangsan” etc, whenever HPAI is first reported in “Gochang”. In fact, by studying the relationship of HPAI occurrence sites in the association rules obtained from the vast amount of data analysis, it would be possible to effectively see potential occurrence sites in advance when the first HPAI disease is reported at a particular site. Further studies into the reason for these associations could be carried out to prevent such a long distance spread from taking place in the future.

3.3. HPAI analysis results using sequential pattern mining

In this section, we are interested in identifying potential sequential dissemination routes of HPAI outbreaks in South

Korea. According to our analysis, we described a process based on a certain scenario that elicits information that generates the sequential dissemination routes of HPAI outbreaks from “Gochang” within a specific period of time. Table 4 presents a sample of the sequential pattern mining rules that were obtained at the city level (support ≥ 0.2) without considering the sites near “Gochang” (to focus on long distance spread only). This result is illustrated in Fig. 8. We can observe that the outbreak is propagated rapidly over a wide area during the first two weeks and spreads more slowly after two weeks; it is propagated throughout the entire nation within six weeks (Jeong et al., 2014; Yoon et al., 2015). Therefore, we can state that it is critical to initiate preventive actions against HPAI in the first two weeks. Also, we depict the sequential propagation routes of HPAI outbreaks from “Gochang” in a GIS version of South Korea (see Fig. 9).

In addition, we can generate the sequential dissemination routes of HPAI outbreaks from the state level to the town level

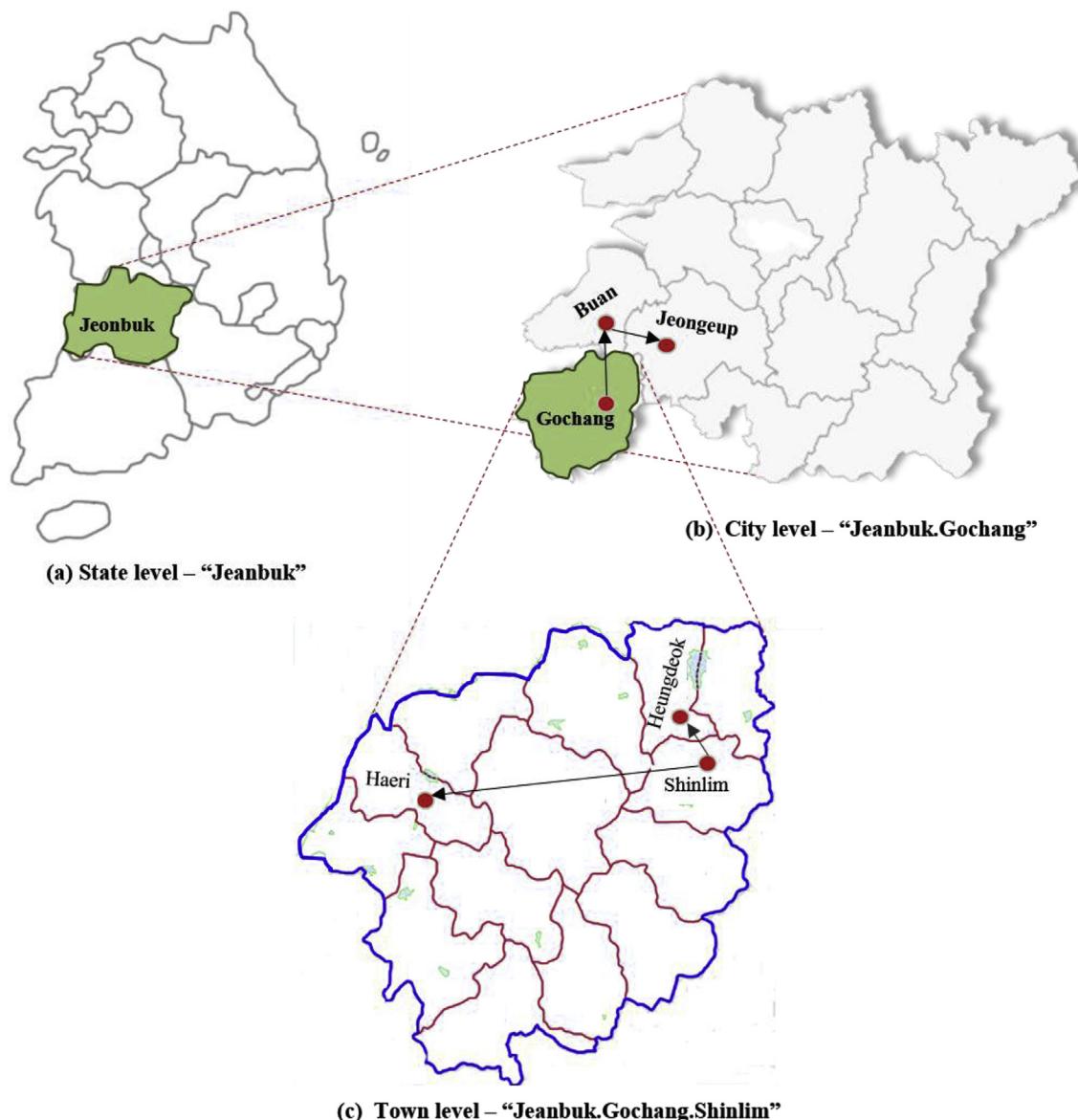


Fig. 11 – Sequential propagation routes of HPAI outbreaks in the location concept hierarchy (from state level to town level).

using the location concept hierarchy (see Fig. 3(d)). Table 5 presents a sample of the obtained sequential pattern mining rules with considering the sites “Gochang.Shinlim” at the town level. Figure 10 shows propagation routes at the town level in the first two weeks. As can be seen in Fig. 10, we can see the HPAI propagation routes from general to specific locations by zooming in and out in the location window. We also depict some of sequential propagation routes of HPAI outbreaks from the state level (“Jeonbuk”) to the town level (“Jeonbuk.Gochang.Shinlim”) with the location concept hierarchy embedded in a GIS representation of South Korea (see Fig. 11).

3.4. Discussions

The system proposed in this paper has two main goals: (1) to organize the data from past outbreaks of HPAI so that it can be described and the data visualized by OLAP, according to various parameters, such as time, location, and so on; and (2) to generate the spatio-temporal spread based on the associations and potential sequential routes derived from past outbreaks by utilizing the data mining technique. Therefore, the main objective of this system is to provide decision makers with the ability to improve the country's preparedness for future outbreaks and to gain a better understanding of how the virus may spread both in space and time over long distances.

It should be noted that the proposal of this system was motivated by the lack of an efficient HPAI data analysis model in order to assist administrators (disease managers and/or epidemiologists involved in the control of these outbreaks) in making more effective decisions. The proposed system is a unique application-driven system. However, it is difficult to declare its effect, as this prototype system has not been verified in the field. Consequently, it is imperative to verify the efficiency of the proposed system through an actual trial; furthermore, the functions of this system should be enhanced accordingly. Finally, we believe that the proposed system can be a useful supplementary tool and an aid for decision makers in the HPAI analysis sessions. In fact, the system itself will not give an answer on the best control measures to use and/or on the processes in place that explain the observed and predicted spread. However, it gives clues to guide further research that, in turn, can provide valuable information about the development of appropriate control measures. Thus, a collaborative research with disease managers and epidemiologists is required in the future.

4. Conclusions

In this paper, we proposed a new design methodology for the analysis of HPAI in South Korea. The core of the methodology is the construction of a data cube model using HPAI occurrence data accumulated in a data warehouse over an extended period. Multidimensional HPAI analysis was performed using OLAP operations, varying the abstraction level according to the temporal and spatial analysis purposes of the HPAI analysis. The proposed analysis model elicited useful information that revealed the site connectedness and the potential sequential dissemination routes of HPAI outbreaks by

applying association rule mining and sequential pattern mining, respectively. We confirmed the feasibility and applicability of the proposed HPAI analysis model by implementing an HPAI analysis system using the Oracle AWM and R tools and then applying it to the analysis of HPAI outbreaks in South Korea. To sum up, our results provide valuable information that can help decision makers to identify situations where outbreaks rapidly spread to distant areas of the country and to further investigate the spread mechanisms responsible for such spreads so that appropriate prevention and control strategies can be developed. Further testing and refinement of the proposed system is required and is a part of our ongoing research.

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-2015R1D1A3A01018731).

REFERENCES

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD international conference on management of data* (vol. 22, pp. 207–216).
- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the 11th international conference on data engineering* (pp. 3–14). IEEE.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed.). Burlington, MA: Morgan Kaufmann Publishers.
- Jeong, J., Kang, H.-M., Lee, E.-K., Song, B.-M., Kwon, Y.-K., Kim, H.-R., et al. (2014). Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its relationship with migratory birds in South Korea during 2014. *Veterinary Microbiology*, 173, 249–257.
- KAHIS (Korean Animal Health Integration System). (2016). *Infectious disease outbreaks in domestic animal*. <http://www.kahis.go.kr/home/lkntscrinfo/selectLkntsOrrorncList.do>. (Accessed 10 may 2016).
- Kim, Y.-S. (2015). Epidemiology and lessons from recent epidemic of highly pathogenic avian influenza (H5N8) in Korea. In *Korean poultry science, 2015 KSPS spring symposium* (pp. 5–38) (in Korean).
- Kim, H.-R., Kwon, Y.-K., Jang, I., Lee, Y.-J., Kang, H.-M., Lee, E.-K., et al. (2015). Pathologic changes in wild birds infected with highly pathogenic avian influenza A (H5N8) viruses, South Korea, 2014. *Emerging Infectious Diseases*, 21(5), 775–780.
- Kwon, J.-H., Lee, D.-H., Swayne, D. E., Noh, J.-Y., Yuk, S.-S., Erdene-Ochir, T.-O., et al. (2016). Highly pathogenic avian influenza A (H5N8) viruses reintroduced into South Korea by migratory waterfowl, 2014–2015. *Emerging Infectious Diseases*, 22(3), 507–510.
- Lee, H.-J., Suh, K., Jung, N.-S., Lee, I.-B., Seo, I.-H., Moon, O.-K., et al. (2014). Prediction of the spread of highly pathogenic avian influenza using a multifactor network: part 2 – comprehensive network analysis with direct/indirect infection route. *Biosystems Engineering*, 118, 115–127.
- MAFRA (Ministry of Agriculture, Food and Rural Affairs). (2016). *Avian influenza information*. (in Korean) http://www.mafra.go.kr/FMD-AI/05/01_05.html. (Accessed 12 September 2016).

- OIE (World Organization for Animal Health). (2016). Avian influenza portal. <http://www.oie.int/animal-health-in-the-world/web-portal-on-avian-influenza/early-detection-warning-diagnostic-confirmation/>. (Accessed 10 may 2016).
- Oracle AWM (analytic workspace manager) package (version 11.2). (2016). <http://www.oracle.com/technetwork/database/options/olap/olap-downloads-098860.html>. (Accessed 10 may 2016).
- R arulesSequences package (version 0.2-6). (2015). <https://cran.r-project.org/web/packages/arulesSequences/index.html>. (Accessed 10 September 2015).
- R arules package (version 1.1-9). (2015). <https://cran.r-project.org/web/packages/arules/index.html>. (Accessed 10 September 2015).
- Seo, I.-H., Lee, I.-B., Moon, O.-K., Jung, N.-S., Lee, H.-J., Hong, S.-W., et al. (2014). Prediction of the spread of highly pathogenic avian influenza using a multifactor network: part 1 – development and application of computational fluid dynamics simulations of airborne dispersion. *Biosystems Engineering*, 121, 160–176.
- Shin, J.-H., Woo, C., Wang, S.-J., Jeong, J., An, I.-J., Hwang, J.-K., et al. (2015). Prevalence of avian influenza virus in wild birds before and after the HPAI H5N8 outbreak in 2014 in South Korea. *Journal of Microbiology*, 53(7), 475–480.
- Tuncer, N., & Le, T. (2014). Effect of air travel on the spread of an avian influenza pandemic to the United States. *International Journal of Critical Infrastructure Protection*, 7(1), 27–47.
- Verhagen, J. H., Herfst, S., & Fouchier, R. A. (2015). How a virus travels the world. *Science*, 347(6222), 616–617. <http://dx.doi.org/10.1126/science.aaa6724>.
- WHO (World Health Organization). (2016). Avian influenza in humans. http://www.who.int/influenza/human_animal_interface/avian_influenza/en/. (Accessed 10 may 2016).
- Yoon, H., Moon, O.-K., Jeong, W., Choi, J., Kang, Y.-M., Ahn, H.-Y., et al. (2015). H5N8 highly pathogenic avian influenza in the Republic of Korea: epidemiology during the first wave, from January through July 2014. *Osong Public Health and Research Perspectives*, 6(2), 106–111.
- Zaki, M. J. (2001). SPADE: an efficient algorithm for mining frequent sequences. *Machine Learning*, 42, 31–60.