19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Interesting-based Association Rules for Highway Traffic Data

Pannawat Sriratna
Department of Computer Science,
Faculty of Science and Technology, Thammasat University
Pathumthani, Thailand
pannavs@tot.co.th

Pakorn Leesutthipornchai
Department of Computer Science,
Faculty of Science and Technology, Thammasat University
Pathumthani, Thailand
pakornl@cs.tu.ac.th

Abstract— **Highway no.7 is the main route for transportation between Bangkok to the eastern seaboard. The eastern seaboard is the major economic region of Thailand. This paper applies theory of rough set to find the hidden relationship among date, time, entering station, exit station and vehicular traffic. The transactional data was gathered from the Division of Inter City Motorway, Department of Highways in the period of January 2014 to April 2015. The set of transactional data consists of 242,066 records. The relations in term of association rules are obtained. The original association results are numerous and difficult to select. This paper proposed the interesting of association rules by using indicators that are generalization, leverage, coverage and lift. The association rules with indicator values are compared and discussed.**

Keywords—Association Rule; Rough Set Theory; Interesting Indicator; Thai Highway Traffic;

I. INTRODUCTION

The highway no.7 is the first international highway of Thailand. It starts from Srinakarin road, Bangkok and ends on Sukhumvit road, Pattaya City, Chonburi Province. Highway no.7 is the main route for transportation between Bangkok to the eastern seaboard. The eastern seaboard is the major economic region of country. It is the tourist attraction and large industrial estates include Pattaya city, Bangsaen beach, Koh Samet, Koh Chang, Amata Nakorn Industrial Estate, Map Ta Phut Industrial Estate and Laem Chabang Port (LCP). LCP is the largest port for import and export the country's products. In addition, the eastern seaboard is connected to three customs that are Aranyaprathet, Chanthaburi, and Khlong Yai Customs. The highway no.7 has been designed as a part of the Asian Highway 19 (AH19) and 123 (AH123). The traffic on the highway no.7 is dense and gets high probability to be increased in the future because the opening of ASEAN Economic Community (AEC).

This paper studies and analyses the traffic data of highway no.7 by applies rough set theory to find the relation between the condition attributes (day, holiday, time, entry toll, exit toll, number of emergency medical responder cars, number of private cars and number of trucks) and the decision attribute (number of vehicles). The data are received from the Division of Inter City Motorway, Department of Highways. The data are recorded from January 2014 to April 2015. The data set consists of 242,066 records with 60,766872 cars.

978-1-4673-7825-3/15/$31.00 ©2015 IEEE

The remainder of this paper is organized as follows. Section II discusses on association mining research works. Section III shows the traffic data set before and after preprocess. Section IV shows rough set association mining process. Section V provides interesting indicators formulation. Section VI discusses on the obtained association results with interesting indicator values. Lastly, section VII concludes our research work.

II. RELATED WORKS

In 2010, Rui Tian and Zhaosheng Yang [1] proposed the method of road traffic accidents causes analysis based on data mining. The research applies association mining to find the relation between factors and traffic accidents. The traffic data are obtained from the ministry of public security. The traffic accident factors are driver, vehicle, road and environmental factors.

In 2011, Sibar Kaan Manga [2] proposed the currency crises prediction with rough set theory. The research applies rough set theory to find the relation among macroeconomic data. The data set is obtained from the web site of central bank of the Republic of Turkey. The economic features are investigated. The examples of economic features are volumes of gold and stock markets, domestic loans and interest rates.

In 2012, Ogwueleka [3] proposed the traffic accident data profiling and clustering with data mining process. The research applies association mining to find the relation between traffic accident features and the injury severity. The data set is collected from the Federal Road Safety Corps and Nigerian Police Force offices. The examples of traffic accident features are vehicle type, driver's age, gender, educational level, vehicle age, accident area, roadway surface condition, light and weather conditions. The target attribute (i.e., injury severity) has four classes that are property damage without injury, slight injury, serious injury and fatal injury.

In 2012, Xiaofeng Zheng and Jianmin Xu [4] presented the studies on the application of rough set analysis in mining of association rules and the realization in provincial road transportation management information system. The research applies association mining to find the relation between transportation features and the quality assessment score. The data set is collected from a highway passenger company of

road transport administration and management information systems in Guangdong Province. The transportation features are transportation security, business behavior, service quality and social responsibility.

In 2014, Caner Erden and Fatih Tüysüz [5] proposed an application of rough sets theory on traffic accidents. The research applies rough set theory to discover the relation between accident features and the number of fatalities in crash. The data set is obtained from official US government website. The accident features are gender, age, race, alcohol, drug involvement, atmospheric condition, first harmful event, injury severity and roadway.

Previous research works applied association mining to find the hidden relationship among the given data. The research work [1-5] applied rough set approach to find relations of traffic data. This paper enhances the benefit of applying rough set theory to find the hidden relationship among incoming features and the number of vehicles.

III. Data Processing

A. Data Set

In this paper, the original traffic data of highway no.7 will be interpreted and rearranged to the record data format before apply to the rough set approach. The original data set is obtained from Division of Inter City Motorway, Department of Highways. The original dataset is stored every hour from January 2014 to April 2015. The example of the dataset is showed in Fig. 1.

From Fig. 1, the set of data is divided into columns that contains toll, hours, entry toll, car type 0 (other cars), car type 1 (4 wheels), car type 2 (6 -10 wheels), car type 3 (more than 10 wheels) and total number of vehicles. The records are separated to two parts that are entry and exit parts. The number in each record displays amount of vehicle passed in an hour period. In this paper, the exit part of data is considered and be transformed to record data as shown in TABLE I.

B. Data Preprocessing

The traffic data in transactional format from Fig. 1 is transformed to record data format as shown in TABLE I. From TABLE I, the columns are date, time, entry toll, exit toll, the number of emergency medical responder (EMR) cars, number of private cars, number of trucks and number of vehicles.

TABLE I. Highway Traffic Data In Record Data Format

Conditon Attributes							Decision
Date	Time	Entry Toll	Exit Toll	EMR	Private	Truck	Vehicle
1/8/14	00:00-00:59	Bangbor	Ladgabang	0	15	26	41
1/8/14	00:00-00:59	Bangpakong	Ladgabang	0	43	14	57
1/8/14	00:00-00:59	Panutnikom	Ladgabang	1	53	35	88
...	...	...	...	...	...	...	...
1/8/14	23:00-23:59	Bangpakong	Ladgabang	0	72	13	85
1/8/14	23:00-23:59	Panutnikom	Ladgabang	0	140	28	168
1/8/14	23:00-23:59	Panthong	Ladgabang	0	717	114	831
...	...	...	...	...	...	...	...
30/4/15	23:00-23:59	Bangbor	Panthong	0	96	4	100
30/4/15	23:00-23:59	Bangpakong	Panthong	0	69	30	99
30/4/15	23:00-23:59	Panutnikom	Panthong	0	52	4	56

Inter City Motorway Division Department of Highways

PCS COMPUTER SYSTEM

Hourly traffic volume, Classified by car type

Date: 01/08/2557 **Destination toll:** Ladkrabang **Channel type:** All
Time: 1 - 24 **Source toll:** All **Channel group:** All
Car type: All **Channel No.:** All

Destination/Hour/Source	Type 0	Type 1	Type 2	Type 3	Total
Entry					
701 - Ladkrabang					
00:00-01:00	0	319	24	139	482
card dispensers	0	318	24	139	481
break through the blockade	0	1	0	0	1
01:00-02:00	0	239	27	86	352
•••		•••		•••	
23:00-00:00	1	726	43	134	904
card dispensers	1	726	43	134	904
All hours	100	40,579	3,023	4,470	48,172
Total Entry	100	40,579	3,023	4,470	48,172
Exit					
701 - Ladkrabang					
00:00-01:00	0	379	46	148	573
701 - Ladkrabang	0	0	1	0	1
free car	0	0	1	0	1
702 - Bangbor	0	15	3	23	41
cash	0	15	3	23	41
703 - Bangpakong	0	43	6	8	57
•••		•••		•••	
23:00-00:00	0	952	54	152	1,158
701 - Ladkrabang	0	1	0	0	1
cash	0	1	0	0	1
702 - Bangbor	0	22	7	35	64
cash	0	22	7	35	64
703 - Bangpakong	0	72	5	8	85
cash	0	72	5	8	85
704 - Panutnikom	0	140	13	17	170
cash	0	140	11	17	168
free car	0	0	2	0	2
705 - Panthong	0	717	29	92	838
cash	0	717	22	92	831
free car	0	0	7	0	7
All hours	0	40,898	3,008	4,655	48,561
Total Exit	0	40,898	3,008	4,655	48,561
All Total	**100**	**81,477**	**6,031**	**9,125**	**96,733**

Fig. 1. The example of original dataset

Since each attribute value in TABLE I has a large number of distinct values. Therefore this paper reduces a large number of distinct values by using a standard data mining technique (i.e., discretization). Each data record is divided into the ranges as shown in TABLE II. After discretization process, we obtained the preprocessed data set in record data format as shown in TABLE III. The condition attributes are day, holiday, time, entry toll, exit toll, the number of EMR cars, number of private cars and number of trucks and decision attribute is the number of vehicles.

TABLE II. The Ranges of Attributes

Attributes		Possible Ranges
Day	Day	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
	Name	Sun, Mon, Tue, Wed, Thu, Fri, Sat
Holiday	Holiday	Yes, No
	Name	Yes, No
Time	Time	{00:00-02:59},{03:00-05:59},{06:00-08:59}, {09:00-11:59},{12:00-14:59},{15:00-17:59}, {18:00-20:59},{21:00-23:59}
	Name	{T1},{T2},{T3},{T4},{T5},{T6},{T7},{T8}
Entry	Entry Toll	Ladkrabang, Bangbor, Bangpakong, Panutnikom, Panthong
	Name	LB, BB, BK, PN, PT

Exit	Exit Toll	Ladkrabang, Bangbor, Bangpakong, Panutnikom, Panthong
	Name	LB, BB, BK, PN, PT
EMR	No. Of EMR Cars	{0},{>=1}
	Name	No, Yes
Private	No. of Private Cars	{<= 1,953},{1,954-3,906 },{3,907-5,859 }, {5,860-7,812},{>=7,813}
	Name	{R1},{R2},{R3},{R4},{R5}
Truck	No. of Trucks	{<=172},{173-344},{345-516},{517-688},{>=689}
	Name	{R1},{R2},{R3},{R4},{R5}
Vehicles	No. of Vehicles	{<=2,062},{2,063-4,124},{4,125-6,186}, {6,187-8,248},{>=8,249}
	Name	{R1},{R2},{R3},{R4},{R5}

TABLE III. THE EXAMPLE OF PREPROCESSED DATA

Condition Attributes								Decision
Day	Holiday	Time	Entry	Exit	EMR	Private	Trucks	Vehicles
Fri	No	T1	BB	LB	No	R1	R1	R1
Fri	No	T1	BK	LB	No	R1	R1	R1
Fri	No	T1	PN	LB	Yes	R1	R1	R1
...	...	...	...	...	...	...	...	...
Fri	No	T8	BB	LB	No	R1	R1	R1
Fri	No	T8	PN	LB	No	R1	R1	R1
Fri	No	T8	PT	LB	No	R1	R1	R1
...	...	...	...	...	...	...	...	...
Thu	No	T8	BB	PT	No	R1	R1	R1
Thu	No	T8	BK	PT	No	R1	R1	R1
Thu	No	T8	PN	PT	No	R1	R1	R1

For example, from TABLE I, the first record is Friday 1st August 2014 that is not a holiday (the value is No). The time is between 00:00 and 00:59 that is T1. The entry toll is Bangbor (BB). The exit toll is Ladkrabang (LB). The number of EMR cars is 0 that is No. The number of private cars is 15 that is in the range R1 (less than 1,953). The number of trucks is 26 that is in the range R1 (less than 172). The number of vehicles is 41 that is in the range R1 (less than 2,062). The overview of data preprocessing is illustrated in Fig. 2.

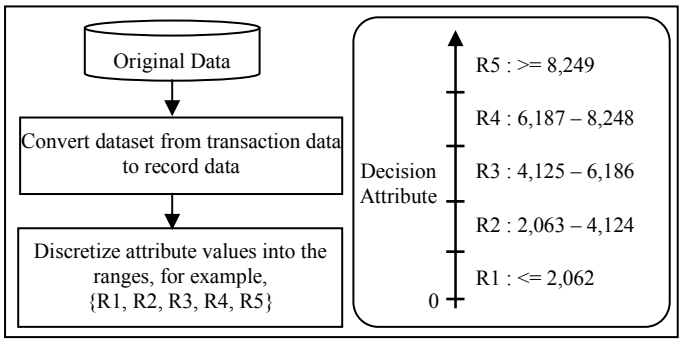

Fig. 2. The overview of data preprocessing

IV. ROUGH SET ASSOCIATION MINING

A. Rough Set Theory: Basic Concept

Rough set theory was first introduced by Pawlak [6]. He applied the rough set theory to knowledge discovery systems. Rough set approach is used to increase the likelihood of correct predictions by removing redundant examples. The applications of rough set theory are characterizing the inter-data relationships, patterns discovering, data analysis and decision making.

The infrequent and frequent relations are found by using the rough set approach. For example, the weather of 100 days is sunny 80 days, cloudy 10 days and rainy 10 days. Every time it rains usually be windy. The rainy day is occurred only 10 percent but every time it rains it always winds. The relationships between two or more attributes in the rainy day subset are found and showed as association rules.

Rough set theory is proposed as a mathematical analysis of incomplete and imprecise knowledge. The basic concept [7] will be described as follows.

1) Information System

Let information system (IS) is the set of records in data set. Each record is considered as an example. Multiple records represent multiple examples. The columns represent attributes for each record. IS = (U, A) where U is finite set of records, $U=\{x_1, x_2, x_3, \ldots, x_n\}$ and A is finite set of attributes. The attributes in set A are classified into two disjoint sets that are condition attributes C and decision attributes D such that $A=C \cup D$ and $C \cap D = \emptyset$.

2) Indiscernibility Relation

Indiscernibility relation is the relation between two or more records where all values are identical in relation to subset of considered attributes. The indiscernibility relation is defined as $R(B)=\{(x,y)\ U\times U : \text{for all } a \in B,\ a(x) = a(y)\}$ where $a \in A$ and $B \subseteq A$.

3) Lower and Upper Approximations

Lower and upper approximations are used to calculate upper and lower bounds. Let $[x]_B$ denotes the equivalence class of B containing x. For any element x of U, $B \subseteq A$ and $X \subseteq U$, the lower approximation of the set X is defined as $\underline{B}X$ and the upper approximation of the set X is defined as $\overline{B}X$. The lower and upper approximations are showed in Eq. (1) and Eq. (2) respectively.

$$\underline{B}X = \{x \in U : [x]_B \subseteq X\} \quad (1)$$

$$\overline{B}X = \{x \in U : [x]_B \cap X \neq \phi\} \quad (2)$$

For given $B \subseteq A$ and $X \subseteq U$, the boundary of X is defined as B(X) as shown in Eq. (3).

$$B(X) = \overline{B}X - \underline{B}X \quad (3)$$

B(X) consists of records that do not certainly belong to X on the basis of A.

The set X which is in the lower approximation will be used to create exact association rules with 100% confidence value and the set X which is in the upper approximation will be used to create "more often association rules" with <100% confidence value.

4) Association Rule

Association rule is created by combining the set of attributes in the rough set. Multiple records are verified an association rule. The decision is taken when condition are indicated by condition attributes are fulfilled. The association rule is presented as implication relation (if…then…).

Let X→Y is the association rule. X is antecedent (left hand side: LHS) set of the rule and Y is consequent (right hand side: RHS) set of the rule. The example of association rule is {Time=T2 and Exit=PT}→{No. of Vehicles=R1} that means if the time is in T2 period (03:00-05:59) at Panthong exit toll then the number of vehicles is less than 2,062 cars.

B. Rough Set Association Mining Procedure

For finding the relationship among day, holiday, time, entry toll, exit toll, number of EMR cars, number of private cars, number of trucks and the number of vehicles, we used rough set theory. Eight condition attributes and one decision attribute are applied to the rough set association mining technique then we obtained the set of association rules which is the hidden relationship of traffic data. The example of results from rough set approach is showed in Fig. 3.

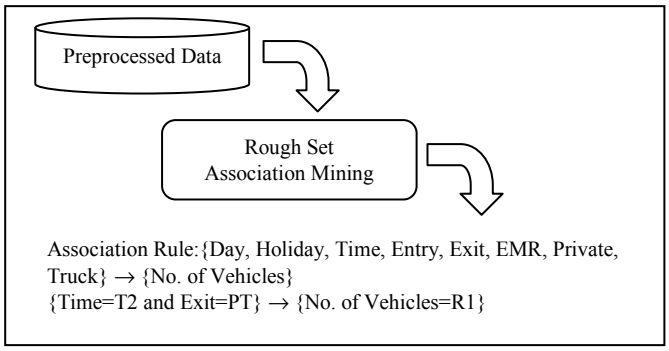

Fig. 3. Rough set association mining

V. INTERESTING INDICATORS

The interesting indicators are described in detail in [8]. This paper focuses on the basic measures that are confidence, support, coverage, leverage, lift and generalization. The descriptions of these measures are showed as followings:

Let $T=\{t_1, t_2, ..., t_n\}$ be the set of transaction records. X→Y is the association rule. X and Y are the item sets where $X \cap Y=\emptyset$. X is antecedent (left hand side: LHS) of the rule and Y is consequent (right hand side: RHS) of the rule.

A. Support

Support is the proportion of transactions which contains the item set X as shown in Eq. (4).

$$\sup(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|} \tag{4}$$

B. Confidence

Confidence is the probability of the support of item set that contains antecedent X and consequent Y over the support of item set X as shown in Eq. (5).

$$\sup(X \rightarrow Y) = \frac{\sup(X \rightarrow Y)}{\sup(X)} = \frac{\sup(X \cup Y)}{\sup(X)} \tag{5}$$

C. Coverage

Coverage is the antecedent (X) support. The coverage indicates that how often does the association rule X→Y is applied. The coverage formulation is showed in Eq. (6).

$$\mathrm{cov}(X \rightarrow Y) = \sup(Y) \tag{6}$$

D. Leverage

Leverage is the difference between the support that X and Y appear together and the support that X and Y appear independently. Leverage indicates that how many items X and Y are sold together more over than the independent sells. The leverage formulation is showed in Eq. (7).

$$\begin{aligned} \mathrm{lev}(X \rightarrow Y) &= \sup(X \rightarrow Y) - \sup(\sup(X) \cdot \sup(Y)) \\ &= \sup(X \cup Y) - (\sup(X) \cdot \sup(Y)) \end{aligned} \tag{7}$$

E. Lift

Lift is the proportion of the support that X and Y appear together over the support that X and Y appear independently. Lift measures that how many times do items X and Y are sold together over than the independent sells. The lift formulation is showed in Eq. (8).

$$\begin{aligned} \mathrm{lift}(X \rightarrow Y) &= \frac{\mathrm{con}(X \rightarrow Y)}{\sup(Y)} \\ &= \frac{\sup(X \cup Y)}{\sup(X) \cdot \sup(Y)} \end{aligned} \tag{8}$$

F. Generalization

Generalization is a measure to indicate the generalization of the association rules. Let N is the number of condition attributes. X is the LHS set of the rule. |X| is the cardinality of the set. The rule with high generalization value is preferred. The generalization formulation is showed in Eq. (9)

$$\mathrm{gen}(X \rightarrow Y) = 1 - \frac{|X|}{N} \tag{9}$$

VI. ASSOCIATION RULES WITH INDICATOR VALUES

Four interesting indicators from section V that are generalization, leverage, coverage and lift are normalized in to the range 0 to 1. Then the summation of indicator values is calculated and showed as interesting score. The total number of association rules obtained from rough set approach is 984. The number of records and rules are showed in TABLE IV. The example of top-10 association rules with highest interesting score are showed in TABLES V-VII. Note that interesting scores of association rules are proposed to select the top-*k* interesting rules from a large number of association results.

This paper considered that the minimum confidence is 100%. The association rules with 100% confidence and high interesting indicator values are preferred.

For interesting indicators, generalization and lift indicators are promoted the rare item set. Rare item set may appear infrequent but every time the condition items appear, the decision item always happens. The leverage and coverage indicators are promoted the popular item set. Popular item set

often appear in most data records. Popular item set is the majority items that have the same set of attributes.

In this paper, we consider three situations that are rare item, popular item, and combining both rare and popular item situations. The top-10 association rules for rare item, popular item and both rare and popular item considerations are showed in TABLES V-VII respectively.

From TABLE V, the sets of rare item rules have high generalization and lift indicator values. The percentage of the rules with R5 decision is only 3.76% from the obtained association rules but they have highest interesting score. The association rule with interesting indicators selects R5 records even if it appears infrequent. The rule no.1 means very large number of private cars and vehicles on Wednesday.

From TABLE VI, the sets of popular item rules have high leverage and coverage indicator values. The percentage of the rules with R1 decision is 41.26% from the obtained association results. The rules that have R1 as decision are often appear. The rule no.7 means few number of cars on Friday in the time between 12:00 and 14:59 passes from Bangpakong toll to Ladkrabang toll (from Chachoengsao to Bangkok).

From TABLE VII, both rare and pupolar item rules are considered together. The first association is few number of vehicles enter from Bangbor toll in every day between 00:00 and 02:59. The rule no.4 means very large number of cars on Saturday in the time between 06:00 and 08:59 is more likely to have an accident.

TABLE IV. THE NUMBER OF DISCOVERED ASSOCIATION RULES

	No. of Vehicles (RHS of the rule)				
	R1	R2	R3	R4	R5
No. of records after preprocess	3,404	394	333	270	104
No. of rules	406	209	190	142	37

TABLE V. ASSOCIATION RULES WITH INTERESTING SCORE FOR RARE ITEM CONSIDERATION

#	Condition	Decision	Con	Gen	Lift	Interesting Score
1	Day=Wed and Private=R5	R5	1.000	0.7500	1.0000	1.7500
2	Day=Thu and Holiday=Yes and Truck=R5	R5	1.000	0.6250	1.0000	1.6250
3	Day=Sat and Time=T3 and Private=R4 and EMR=Yes	R5	1.000	0.5000	1.0000	1.5000
4	Day=Fri and Holiday=Yes and Time=T6 and Truck=R5	R5	1.000	0.5000	1.0000	1.5000
5	Holiday=Yes and Time=T6 And Truck=R4 and EMR=Yes	R5	1.000	0.5000	1.0000	1.5000
6	Day=Sat and Time=T4 and Entry=LB and Truck=R3 and EMR=Yes	R5	1.000	0.3750	1.0000	1.3750
7	Day=Fri and Holiday=Yes and Time=T6 and Entry=LB and Private=R4 and Truck=R4	R5	1.000	0.2500	1.0000	1.2500
8	Entry=BK and Private=R3	R3	1.000	0.7500	0.3123	1.0623
9	Time=T1 and Truck=R5	R2	1.000	0.7500	0.2640	1.0140
10	Day=Sat and Time=T8 and Truck=R4	R4	1.000	0.6250	0.3852	1.0102

TABLE VI. ASSOCIATION RULES WITH INTERESTING SCORE FOR POPULAR ITEM CONSIDERATION

#	Condition	Decision	Con	Lev	Cov	Interesting Score
1	Day=Thu and Time=T1 and Truck=R3	R1	1.0000	0.9998	0.7556	1.7554
2	Day=Wed and Time=T8 and Truck=R2 and EMR=Yes	R1	1.0000	0.9998	0.7556	1.7554
3	Day=Thu and Time=T2 and Truck=R3	R1	1.0000	0.9995	0.7556	1.7551
4	Day=Tue and Private=R1 and Truck=R3 and EMR=Yes	R1	1.0000	0.9995	0.7556	1.7551
5	Exit=BB and Truck=R5	R1	1.0000	0.9993	0.7556	1.7549
6	Day=Wed and Time=T2 and Truck=R3	R1	1.0000	0.9993	0.7556	1.7549
7	Day=Fri and Time=T5 and Entry=BK and Exit=LB	R1	1.0000	0.9993	0.7556	1.7549
8	Day=Fri and Time=T2 and Truck=R3	R1	1.0000	0.9992	0.7556	1.7548
9	Day=Tue and Time=T2 and Truck=R3	R1	1.0000	0.9992	0.7556	1.7548
10	Day=Wed and Holiday=Yes and Time=T5 and Exit=PN	R1	1.0000	0.9992	0.7556	1.7548

TABLE VII. ASSOCIATION RULES WITH INTERESTING SCORE

#	Condition	Decision	Con	Gen	Lev	Cov	Lift	Interesting Score
1	Time=T1 and Entry=BB	R1	1.0000	0.7500	0.9878	0.7556	0.0004	2.4938
2	Day=Wed and Private=R5	R5	1.0000	0.7500	1.0000	0.0231	1.0000	2.7731
3	Day=Thu and Holiday=Yes and Truck=R5	R5	1.0000	0.6250	1.0000	0.0231	1.0000	2.6481
4	Day=Sat and Time=T3 and Private=R4 and EMR=Yes	R5	1.0000	0.5000	1.0000	0.0231	1.0000	2.5231
5	Day=Fri and Holiday=Yes and Time=T6 and Truck=R5	R5	1.0000	0.5000	1.0000	0.0231	1.0000	2.5231
6	Holiday=Yes and Time=T6 and Truck=R4 and EMR=Yes	R5	1.0000	0.5000	1.0000	0.0231	1.0000	2.5231
7	Exit=BB and Truck=R5	R1	1.0000	0.7500	0.9993	0.7556	0.0076	2.5126
8	Exit=BB and EMR=Yes	R1	1.0000	0.7500	0.9940	0.7556	0.0008	2.5004
9	Time=T2 and EMR=Yes	R1	1.0000	0.7500	0.9930	0.7556	0.0007	2.4993
10	Time=T1 and EMR=Yes	R1	1.0000	0.7500	0.9919	0.7556	0.0006	2.4982

VII. CONCLUSION

This paper applied rough set theory to find the hidden relation of traffic on highway no.7 of Thailand. Traffic data are obtained from Division of Inter City Motorway, Department of Highways from January 2014 to April 2015. This paper defines the condition attributes that are day, holiday, time, entry toll, exit toll, the number of EMR, number of private cars and number of trucks. The decision attribute is the number of vehicles. The association rules with 100% minimum confidence constraint are considered. The results

show that the number of vehicles is related to day, holiday, time, entry toll, exit toll, the number of EMR, number of private cars and number of trucks.

This paper proposed the association rule with interesting indicator values. Interesting indicators are generalization, leverage, coverage and lift. The top-10 association rules with 100% confidence and highest indicator values are selected from 984 rules.

REFERENCES

[1] R. Tian and Z. Yang, "Method of Road Traffic Accidents Causes Analysis Based on Data Mining", Proc. of International Conference on Computational Intel-ligence and Software Engineering (CiSE), 2010, pp. 1-4.

[2] S. K. Manga, "Currency Crises Prediction with Rough Set Theory", International Journal of Computer Applications, vol. 32, 2011, pp. 48-52.

[3] F. N. Ogwueleka and T. C. Ogwueleka, "Traffic Accident Data Profiling and Clusteringwith Data Mining Process", IOSR Journal of Computer Engineering, vol.6, 2012, pp. 14-22.

[4] X. Zheng and J. Xu, "Studies on the Application of Rough set Analysis in Mining of Association Rules and the Realization in Provincial Road Transportation Management Information System", International Conference on Industrial Control and Electronics Engineering, 2012.

[5] C. Erden and F. Tüysüz, "An Application of Rough Sets Theory on Traffic Accidents", Proc. of International Conference on Engineering and Applied Sciences Optimization, 2014.

[6] Z. Pawlak, Rough Sets, "In Theoretical Aspects of Reasoning about Data", Kluwer, Netherlands, 1991.

[7] M N. A Rahman and Y. M Lazim and F. Mohamed, "Applying Rough Set Theory in Multimedia Data Classification", The Society of Digital Information and Wireless Communications, 2011.

[8] L. Geng and HJ. Hamilton, " Interestingness Measures for Data Mining: A Survey", ACM Computing Surveys, vol. 38, 2006.

[9] Z. Pawlak and J. Grzymala-Busse and R. Slowinski and W. Ziarko, "Rough Sets", Communications of the ACM, vol. 38, 1995, pp. 88-95.

[10] J. Hong-bo and Y. De-li, "Application Research on Fast Discovery of Association Rules Based on Air Transportation", Proc. of International Conference on Service Systems and Service Management, 2007, pp. 1-6.

[11] M H. F. Zarandi, "Application of Rough Set Theory in Data Mining for Decision Support Systems (DSSs)", Journal of Industrial Engineering, vol.1, 2008, pp.25-34.

[12] Y. Zhao and H. Jiang and R. Geng and X. Dong, "Mining Weighted Negative Association Rules Based on Correlation from Infrequent Items", Advanced Computer Control, 2009. ICACC '09. International Conference on, 2009.