IEICE TRANS. INF. & SYST., VOL.E92–D, NO.1 JANUARY 2009

PAPER

Weighted Association Rule Mining for Item Groups with Different Properties and Risk Assessment for Networked Systems

Jungja KIM[†a)], **Heetaek CEONG**[††], ***Nonmembers***, ***and*** **Yonggwan WON**[†††*], ***Member***

SUMMARY In market-basket analysis, weighted association rule (WAR) discovery can mine the rules that include more beneficial information by reflecting item importance for special products. In the point-of-sale database, each transaction is composed of items with similar properties, and item weights are pre-defined and fixed by a factor such as the profit. However, when items are divided into more than one group and the item importance must be measured independently for each group, traditional weighted association rule discovery cannot be used. To solve this problem, we propose a new weighted association rule mining methodology. The items should be first divided into subgroups according to their properties, and the item importance, i.e. item weight, is defined or calculated only with the items included in the subgroup. Then, transaction weight is measured by appropriately summing the item weights from each subgroup, and the weighted support is computed as the fraction of the transaction weights that contains the candidate items relative to the weight of all transactions. As an example, our proposed methodology is applied to assess the vulnerability to threats of computer systems that provide networked services. Our algorithm provides both quantitative risk-level values and qualitative risk rules for the security assessment of networked computer systems using WAR discovery. Also, it can be widely used for new applications with many data sets in which the data items are distinctly separated.

key words: *weighted association rule, different item groups, networked systems, risk assessment*

1. Introduction

The main objective of the association rule mining is to discover the interesting associations or relevant relations among itemsets in a large volume of data [1], [2]. For example, the rule 'chair → table' (with support = 0.6, confidence = 0.9) says that the probability of buying both a chair and a table is 0.6, and that the probability of buying a table is 0.9 given that a transaction contains a chair (a chair is already bought). The chair and table are similar in their property called *furniture*, and this fact can be discovered from their frequency of occurrence in the transaction database. In traditional association rule mining, all items in a transaction that have the same significance, without taking account of their weights, are treated uniformly [1]–[3]. However, weighted association rule (WAR) mining considers the items independently [4], [5]. For example, a marketing manager may want to sell the more profitable goods which should become more *important* than other items. When he sells a sofa and a chair, if the profit of the sofa is much higher than that of the chair, then the rule 'sofa → table' (with support = 0.6, confidence = 0.7) should be more *important* than the rule 'chair → table' (with support = 0.8, confidence = 0.9), although it has higher values for support and confidence computed by ordinary association rule mining method. Weighted association rule discovery reflects the importance of each item in the point-of-sale database, in contrast to traditional association rule mining, and it means that more profitable items are discovered according to the rules [6]–[8].

In this paper, we apply WAR discovery to network security management [9], [10]. We assume that the manager wants to know the degree to which each networked system is vulnerable to certain threatening factors. Traditional association rule discovery assumes that the tender/threaten factors have uniform impact and it generates rule sets by considering only item frequency, so it cannot provide a satisfactory answer. However, WAR discovery generates the risk rule set by considering item weights according to the importance of particular items (system, service etc.), enabling a more detailed network risk analysis. For example, WAR discovery can reflect the situation that a system, which provides *web* services in a Windows NT environment, is a tender with greater significance than a system which provides *ftp* service in the same Windows NT environment.

WAR discovery assumes that even though an event may take place only once, it has significant impact, and the rules should reflect it. In traditional association rule discovery, because the rules 'Windows NT → web' and 'Windows NT → ftp' are considered with only frequency, those two rules have the same effect. Thus, traditional association rule discovery cannot be used to manage the network effectively from a qualitative point of view. However, in the weighted association rules, each item is calculated by weight, so the fact that *web* services are weaker against hacking attempts can be introduced by the rule that 'Windows NT → web is weaker than Windows NT → ftp with 20% support'. Also, as discovered risk rules are calculated by various weighting factors, detailed numerical operations are possible in network risk management.

When weighted association rule mining has been used in market-basket analysis, each item was attached with a numerical weight given by the users [4], [7]. However, in the case of network operational data [9], the properties of the

Manuscript received February 26, 2008.
Manuscript revised July 15, 2008.

[†]The author is with Chonbuk National University, Chonju, Korea.
[††]The author is with Chonnam National University, Yeosu, Korea.
[†††]The author is with Chonnam National University, Gwangju, Korea.
[*]To whom all correspondence should be sent.
a) E-mail: jungjakim@chonbuk.ac.kr

DOI: 10.1587/transinf.E92.D.10

data items that compose each transaction are different, so that the typical weight association rule methods cannot be applied. To solve this problem, the data items should be grouped according to their properties, and the item weight should be newly defined.

In this paper, we propose a novel methodology that assigns importance to the data that composes different data groups, such as network operational data [9]. In our proposed methodology, the items are first divided into subgroups according to the properties of the items, and their importance, i.e. item weight, is defined or calculated only with the items included in the subgroup. Then, transaction weight is measured by appropriately summing up the item weights that are in turn calculated from each subgroup, and the weighted support is computed as the fraction of the transaction weights that contains the candidate items relative to the weight of all transactions. As a result, our algorithm provides quantitative risk-level values and qualitative risk rules for assessing the risk of networked computer systems against illegal attacks. It also provides network risk analysis models which define risk-level values by newly defined weighting factors.

2. Weighted Association Rule Discovery

Although the discovery of association rules and weighted association rules are similar, there are differences: One is the process that generates candidate itemsets using various weighting factors, (i.e., item weight, transaction weight etc) in each step, while the other is defining large itemsets pruned by minimum weighted support [7], [8], [11].

2.1 Association Rules

An association rule can provide valuable knowledge representation which can represent the implicit correlations among the items in a large number of transactions [3], [12], [13]. Given $I = \{i_1, i_2, \ldots, i_m\}$ as the items' space, which is a set of m distinct items (database attributes), let D is a dataset (database) which is a set of transactions and each transaction T_j be defined as a set of items (itemset; subset of I) such that $T_j \subseteq I$ and $T = D = \{T_1, T_2, \ldots, T_l\}$.

An association rule has the form of $X \rightarrow Y$ where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. Note that X and Y are transactions (itemsets) which can include a single or multiple items. The significance of an association rule is determined by two measurements, *support* and *confidence*. The first measurement *support* is the frequency that the itemsets (X and Y) occur or co-occur in a transaction database D, which can be also considered as the probability that $X \cup Y$ exists in a transaction T_j in the database D. The other measurement *confidence* represents how "strongly" an itemset X implies another itemset Y, which can be also considered as the probability that Y exists given that a transaction contains X. *Support* and *confidence* of the rule $X \rightarrow Y$ can be computed by

- $\text{support}\,(X \rightarrow Y) = \dfrac{n(X \cup Y)}{|D|} = P\,(X \cup Y)$
- $\text{confidence}\,(X \rightarrow Y) = \dfrac{\text{support}\,(X \cup Y)}{\text{support}\,(X)} = \dfrac{n(X \cup Y)}{n(X)} = \dfrac{P\,(X \cup Y)}{P\,(X)} = P\,(Y|X)$

where $n(S)$ is the number of transactions which contain the itemset S, and $|D|$ is the size of the database D.

An itemset is said to be *frequent* if its support is larger than a user-specified threshold value (i.e., minimum support(min_sup)). Also, support($X \rightarrow Y$) larger than min_sup implicates that, when X appears in a transaction, Y is more likely to appear in the same transaction with a *confidence* value confidence($X \rightarrow Y$). Finally, we should note that an association rule Rule($X \rightarrow Y$) exists, if support($X \rightarrow Y$) and confidence($X \rightarrow Y$) are both greater than their corresponding threshold value [1]–[3], [12], [14].

2.2 Weighted Association Rules

The weighted association rule expands the traditional association rule by allowing a weight, which reflects the *importance* of each item in a transaction. The item importance is used to calculate various weighting factors. According to the definition of the association rule, we can define the weighting factors that are applied in the weighted association rule.

The items can be weighted within different weighting spaces depending on the mining focus. Weighting space is the context within which the weights are evaluated. The weighting factors are the values calculated in acceptable weighting spaces. In this paper, WS_t is inner-transaction space that refers to the host transaction that an item is weighted in. WS_I is item space that refers to the space of the item collection that covers all the items appearing in the transactions. WS_T is transaction space that is defined for transactions rather than for items [7]. The weighting factors, definitions and properties related to the weighted association rule are briefly described below [4], [5], [8].

Weight $w(i)$: Given a set of items $I = \{i_1, i_2, \ldots, i_n\}$, we assign a weight $w(i_j)$ for each item i_j, with $0 \leq w(i_j)$, where $j = \{1, 2, \ldots, n\}$. This allows expression of the item importance (i.e., significance). These kinds of weights are the itemset weight and the transaction weight which are described as follows:

1) Itemset Weight $w(is)$: It is based on the item weight $w(i_j)$. The weight of an itemset denoted as $w(is)$ can be derived from the weights of items included in the itemset. One simple way is to calculate the average value of the item weights as described in the equation (1):

$$w(is) = \frac{\sum_{k=1}^{|is|} w(i_k)}{|is|}. \tag{1}$$

2) Transaction Weight $w(t_k)$: It is a type of itemset weight and a value attached to each of the transactions, denoted as the equation (2). It can be derived from weights of the items presented and formulated easily in the average weight of the items presented in the transaction. At this time, $WS_t(t_k)$ denotes the inner-transaction space for the k-th transaction in transaction space WS_T,

$$w(t_k) = \frac{\sum_{i=1}^{|WS_t(t_k)|} weight(item(i))}{|WS_t(t_k)|}. \quad (2)$$

Large Itemset: A k-itemset X is called large itemset if the weighted support of such itemset is greater than or equal to the minimum weighted support threshold.

Important Rule: A weighted association rule $X \rightarrow Y$ is called an *important* rule if $X \cup Y$ is a *large* itemset and the confidence of the rule is greater than or equal to a minimum confidence threshold.

2.3 Algorithm for Weighted Association Rules

The algorithm of weighted association rule is composed of two phases. In first phase, it generates the candidate itemset for the large itemset. This is same as the traditional association rule discovery. Next phase is the step that calculates various weighting factors, which satisfy the definition of weighted association rule. In this phase, it calculates weighted support of each candidate itemsets and decides the large itemsets, which satisfy minimum weighted support among candidate itemsets. This process is repeated until no more rules are generated [5], [7], [8], [11]. Figure 1 shows the pseudo code for the algorithm of weighted association rule discovery.

```
ALGORITHM: Weighted Association Rule
input : Transaction database T, minimum weighted support
(minwsp)
L1 = {large 1-item set};
for (i = 2 ; Li-1≠Ø ; i++) do begin
    Ci = apriori-gen(Li-1);      // New candidate generate//
    for all transactions t ∈ T do begin
        (SC, C1) = computing(T, w);
            //the calculation of improved weighted factors //
            //SC : transaction number including given item//
        Ct = subset(Ci, t);  // Candidates contained in t //
        for all candidates c ∈ Ct do
            c.count ++;
        end
        Li = {c ∈ Ci | c.count ≥ minwsp}
    end
Rules(SC, L) = L ∪ Li;
end
```

Fig. 1 Algorithm for weighted association rule.

3. WAR Discovery for Different Item Groups

In this paper, we apply weighted association rule discovery to the domain of network risk assessment. Network operational data are composed of a data item group, in which each itemset in the transaction has different properties. In these cases, we cannot apply a traditional weighted association rule discovery. Instead, we must consider two major differences compared to traditional weighted association rule discovery. Firstly, it is the definition of item weight. In the point-of-sale database, item weight has a predefined initial value. However, the transaction itemset is composed of a data group with different properties. Therefore, to define the reasonable weights, the weight should be individually calculated in a different data item group. We propose improved weighting factors in Definition 1, Definition 2, and Definition 3 and do not consider confidences in this paper. Secondly, it mines rules by applying newly defined weighting factors in deciding the large itemsets from the candidate itemsets. We discovered the risk rules for network operational data, and defined the minimum weighted support that is calculated in rule discovery as the risk-level value.

3.1 Risk Assessment of Networked Systems

Table 1 shows the symbols and definitions used in our proposed model, and Table 2 shows the transaction database. The vulnerability/threat database supposes that each item is composed of system (OS), services and risk value as shown in Table 2. The risk value indicates how a system with a transaction item will be critical for providing the services if the system fails. It can be determined by domain experts, system managers or official organizations such as KISA (Korea Information Security Agency).

It is obvious that a transaction item which is more *frequent* in the vulnerability/threat database has been more exposed to risky situations such as hacking attempts and should have a higher weight value. As usual, the item weight is a predefined value reflecting the characteristics of the domain. However, we should define weight with regard to the fact that the characteristics of the items composing the transaction are different. As shown in Table 2, the items in transaction are composed of several services and one OS. So,

Table 1 Symbols and definitions.

Symbol	Definition
w	Itemset weight
T	Transaction database
T(o,s)	Each transaction item consists of system(o) and services(s)
O	Operating System(OS) itemset
S	Service itemset
ic	Item frequency
rv	Risk value
nrv	Normalized risk value

Table 2 Transaction list and risk value of vulnerability/threat database. Transaction weight is calculated by the equation (2).

ID	T(system(OS), service)	Risk val. (rv)	Trans. wts
T1	(Window 2000, WEB, DNS)	5	1.85
T2	(Linux 7.1, FTP, DNS)	3	1.75
T3	(Window 2000, RPC, SMTP, DNS)	3	1.68
T4	(Solaris 8, WEB, DNS, FTP, Telnet)	5	1.82
T5	(Linux 7.1 WEB, DNS, FTP, Telnet, SMTP)	3	1.73
T6	(Solaris 8, WEB)	1	1.92
T7	(Windows 2000, DNS, Telnet)	3	1.76
T8	(Solaris 8, WEB, DNS, Telnet, SMTP)	3	1.79
T9	(Linux 7.1, WEB, DNS)	3	1.82
T10	(Window NT, SMTP, DNS)	3	1.59
T11	(Solaris 8, WEB, FTP, Telnet)	5	1.81
T12	(Solaris 8, WEB, Telnet, RPC)	3	1.73
T13	(Window NT, DNS, FTP, Telnet)	1	1.63
T14	(Solaris 8, WEB, FTP, SMTP)	3	1.79
T15	(Windows 2000, WEB, FTP, RPC, SMTP)	3	1.7
T16	(Linux 7.1, WEB, FTP)	5	1.8
T17	(Window NT, WEB, DNS, FTP)	3	1.69
T18	(Solaris 8, FTP, RPC, DNS)	1	1.71
T19	(Windows 2000, WEB)	3	1.9
T20	(Window NT, WEB, DNS)	5	1.72

when we calculate the frequency or the weight, it should be defined with different criteria.

3.2 Improved Weighting Factors according to Item Group

The transaction itemsets in Table 2 are categorized in two groups. The one is system group and the other is service group. For example, Solaris 8, Linux 7.1, Windows NT are system group and Web, DNS, telnet, SMTP are service group. They have different properties. It is composed of multiple services to one system. So, to define reasonable item weight, it should be calculated to system group and services group distinctly.

The item weight of the proposed model is shown in Definition 1. The item weight is defined as sum of item support and item significance (i.e., importance). The item support is defined as the sum of the item frequency in one item group. The item significance is the sum of the risk values of the transaction divided by the item-count. In our proposed model, nrv is the parameter for normalization to balance with a frequency rate between 0 and 1.

Definition 1: The system $O(i) = \{o_1, o_2, \ldots o_n\}(1 \le i \le n)$ and the service $S(j) = \{s_1, s_2, \ldots s_m\}(1 \le j \le m)$ are in transaction database (T). Item weight (wi) represents that system weight $w(o_i)$ and service weight $w(s_j)$, as defined in the equations (3), (4) and (5). This definition means that each item frequency and the item weight is defined or calculated only with the items included in the subgroup

$$wi = \textit{Item support} + \textit{Significance(rv)}, \; nrv \in [0, 1] \quad (3)$$

Table 3 Item frequency, item support, significance, and item weight.

Item	Item freq.(ic)	Item support	Signif-icance	Item weight
Solaris 8	7	0.35	0.6	0.95
Win. 2000	5	0.25	0.68	0.93
Win. NT	4	0.2	0.6	0.8
Linux 7.1	4	0.2	0.7	0.9
WEB	14	0.26	0.71	0.97
DNS	13	0.24	0.63	0.87
FTP	10	0.19	0.64	0.83
Telnet	7	0.13	0.66	0.79
RPC	4	0.07	0.5	0.57
SMTP	6	0.11	0.6	0.71

$$w(o_i) = \frac{n(o_i)}{n(T(O))} + \alpha \frac{\sum rv(o_i)}{ic(o_i) \times nrv} \quad (4)$$

$$w(s_j) = \frac{n\left(s_j\right)}{n(T(S))} + \alpha \frac{\sum rv\left(s_j\right)}{ic\left(s_j\right) \times nrv} \quad (5)$$

where $n(\cdot)$ represents the number of elements.

In equations (4) and (5), α is the user defined value for relating significance to frequency. In this paper, when calculating the weight for service $w(s_j)$, we defined it as 1 because the relationship 'frequency $\ll$ risk' value is generally made. And, it is reasonable that the degree of risk to a provided service or OS is emphasized more than the frequency. nrv is normalized risk value which can be computed by $rv(Ti)/5$ because maximum rv is 5 as shown in Table 2.

For example, the item frequency of 'solaris 8' is 7 as shown in Table 3. The sum of item frequency in system group is 20 and that in service group is 54. The item support of 'solaris 8' (7 divided by 20) is 0.35, while the significance of 'solaris 8' is calculated as ((5+1+3+5+3+3+1)/(7×5) = 0.6). Therefore, item weight $w(s_j$: j = solaris 8) = 0.35 + 0.6 = 0.95. The calculated results of weighting factors for all items in the database are shown in Table 3.

Definition 2: Transaction weight $w(t_k)$ is measured by appropriately summing the item weights each of which is calculated from its subgroup, as denoted in the equation (6).

$$w(t_k) = \frac{\sum_i w(o_i)}{n(o_i)} + \frac{\sum_j w\left(s_j\right)}{n\left(s_j\right)} \quad (6)$$

where $n(\cdot)$ represents the number of elements.

For example, $w(T1)$ is calculated as follows. According to the Definition 1, the item weight for the system group having a single item w(Windows 2000) is 0.93, and that for the service group having double items w(WEB, DNS) is computed by (0.97 + 0.87)/2 = 0.92. Therefore, $w(T1)$ = 0.93 + 0.92 = 1.85. Each transaction weight is calculated in Table 2.

Definition 3: Weighted support of an itemset(wsp). A set of transactions T respects a rule R in the form $X \rightarrow Y$,

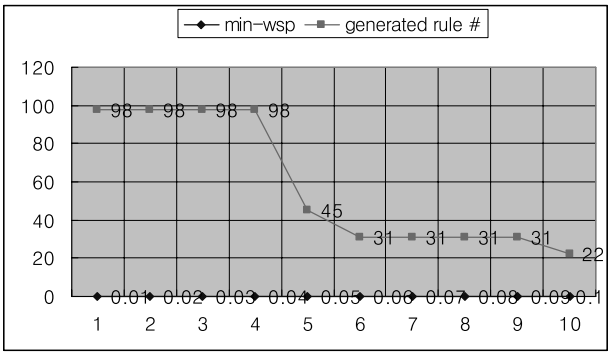

Fig. 2 The generated rule numbers according to min_wsp.

where X and Y are finite sub-itemsets of the item space I and they share no item in common. The *wsp* is computed as the fraction of the transaction weights, where the transaction contains the candidate items, relative to the weight of all transactions. It is used as statistic value for pruning large items. In the weighted association rule, the pruned large item sets satisfy the 'weighted downward closure property' which means that when a certain item set is large, the subset is also large [4], [7]. This can be formulated as the equation (7).

$$wsp(XY) = \frac{\sum_{k=1}^{|WS_T|(X\cup Y)\subseteq t_k} w(t_k)}{\sum_{k=1}^{|WS_T|} w(t_k)}. \quad (7)$$

For example, item set (Solaris 8, DNS) appears in transaction 4, 8 and 18 as shown in Table 2. The sum of the transaction weights is 35.19, therefore *wsp*(Solaris 8, DNS) = (1.82 + 1.79 + 1.71)/35.19 = 0.15

4. Experiments and Analysis

Our proposed methodology was applied to assess the vulnerability/threat for computer systems that provide networked services. According to the algorithm of the weighted association rule, we created the vulnerability/threat rules based on the large items, which satisfies the 'weighted downward closure property' by exceeding the minimal weighted support(min_wsp). The *wsp* was defined in mined rules as the statistical value which defines important item sets generated in each transaction, and it is considered as the risk-level value of the vulnerability/threat.

As a result, our algorithm provides a quantitative risk-level value for the security assessment of the computer systems. In this paper, we consider the experimental data shown in Table 2 with min_wsp from 0.01 to 0.1. Figure 2 shows the number of generated rules according to min_wsp. Table 4 shows the 22 risk rules discovered with min_wsp = 0.1. For example, it shows that Linux7.1 from the rule R4 carries a risk of 0.15 relative to FTP, and that the most vulnerable service to Solaris 8 is WEB which has risk value 0.3 in the discovered rules R1, R2, R3, R6, and R8. Thus, the discovered rules provide more useful information for network risk assessment. We think that it is not meaningful to compare this approach to previous research in performance evaluation. The reason is that this proposed approach is a novel method for grouping items with different properties, which is distinct from previous WAR discovery approaches that treat items with same proprieties equally.

Table 4 Discovered risk rules.

Rule No.	Risk rule	*wsp* (risk-level value)
R1	Solaris 8 →Telnet	0.20
R2	Solaris 8 → SMTP	0.10
R3	Solaris 8 → FTP	0.20
R4	Linux 7.1 → FTP	0.15
R5	Window NT → DNS	0.19
R6	Solaris 8 → DNS	0.15
R7	Linux 7.1 → DNS	0.15
R8	Solaris 8 → WEB	0.30
R9	Linux 7.1 → WEB	0.15
.	.	.
.	.	.
.	.	.
R19	Solaris 8 → Telnet, WEB	0.20
R20	Solaris 8 → Telnet, FTP	0.10
R21	Solaris 8 → FTP, WEB, Telnet	0.10
R22	Solaris 8 → DNS, WEB, Telnet	0.10

5. Conclusions

Weighted association rules are a generalization of traditional association rules in the sense that they allow to use the *importance* (or weight) of items in the transactions. In previous studies of the management of a point-of-sale database, weighted association rule discovery resulted in important rules by reflecting initially defined weight. However, when the properties of data items that make up each transaction are different, we cannot directly apply traditional weighted association rule discovery.

To solve this problem, this paper proposes a new method to discover weighted association rules. The items should be first divided into subgroups according to the properties of the items, and the item weight is defined or calculated only with the items included in the subgroup. To evaluate our rule generation method for the groups with different properties, we applied it to vulnerability assessment for networked computer systems, and defined the level of risk rules using newly calculated weighting factors.

The major contribution of our proposed approach is twofold. In first, it proposes a new methodology to discover weighted association rules, for the case where items are composed of different subgroups with different weight measures. Thus, it can be widely used for applications in which the data items are distinctly separated into groups with different properties. Secondly, it applies the weighted association rule problem to the new application domain of network risk assessment. Even though we used artificially generated

data set to provide better understanding of our method, with a real-world data set, it can be applied effectively to wide range of communication network management [9], [10].

Acknowledgement

"This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MEST, KOSEF)" (The Regional Research Universities Program/Center for Healthcare technology Development, No. R01-2007-000-20926-0).

References

[1] A. Savasere, E. Omiencinsky, and S. Navathe, "An efficient algorithm for mining association rules in large databases," Proc. 21st Int'l Conf. on Very Large Data Bases(VLDB'95), pp.432–444, Zurich, Swizerland, 1995.
[2] R. SriKant and R. Agrawal, "Mining generalized association rules," Proc. 21st Int'l. Conf. on Very Large Data Bases(VLDB'95), pp.407–419, 1995.
[3] E.-H. Han, G. Karypis, and V. Kumar, "Scalable parallel data mining for association rules," Proc. ACM SIGMOD, pp.277–288, Tucson, U.S.A., 1997.
[4] F. Tao, Mining binary relationships from transaction data in weighted settings, PhD Thesis, School of Computer Science, Queen's University Belfast, UK, 2003.
[5] W. Wang, J. Yang, and P. Yu, "Efficient mining of weighted association rules(WAR)," Proc. ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, pp.270–274, 2000.
[6] D.L. Olson, "Mining fuzzy weighted association rules," Proc. 40th IEEE International Conference on System Sciences, 2007.
[7] F. Tao, F. Murtagh, and M. Farid, "Weighted association rule mining using weighted support and significance framework," Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.661–666, 2003.
[8] C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong, "Mining association rules with weighted items," Proc. 1998 International Symposium on Database Engineering & Applications, pp.68–77, 1998.
[9] A. Clemm, Network Management Fundamentals, Cisco Press, 2006.
[10] C. McNab, Network Security Assessment: Know Your Network, O'Reilly Media, 2004.
[11] G.D. Ramkumar, S. Ranka, and S. Tsur, "Weighted association rules: model and algorithm," Proc. Fourth International Conference on Knowledge Discovery and Data Mining, New York City, Aug. 1998.
[12] J. Han and Y. Fu, "Discovery of multiple-level association rules from large databases," Proc. 21st Int'l Conf. on Very Large Data Bases(VLDB'95), pp.420–431, Zurich, Switzerland, 1995.
[13] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Efficient mining of association rules using closed itemset lattices," Information Systems, vol.24, no.1, pp.25–46, 1999.
[14] Y.-C. Lee, T.-P. Hong, and W.-Y. Lin, "Mining association rules with multiple minimum supports using maximum constraints," Int'l. J. Approx. Reason., vol.40, pp.44–54, 2005.

Jungja Kim received the B.S., M.S. degree in 1985, 1988 and Ph.D. degree in from 1997 to 2002, in Computer Science from Chonnam National University respectively. She worked with electronic telecommunication Laboratory at Chonnam National University from 2002 to 2004, and Korea Bio-IT Foundry Center at Gwangju from 2004 to 2006. She is currently an assistant professor at Chonbuk National University. Her major research interest is the bio and medical data analysis, database security, and pattern recognition.

Heetaek Ceong received the B.S., M.S. and Ph.D. degree in Computer Science from Chonnam National University in 1992, 1995 and 1999, respectively. He is currently an associate professor in Chonnam National University in Korea, and director of Information Science Research Center at Yeosu. His major research interest is the bio data analysis, RFID/USN system, and Multimedia.

Yonggwan Won received the B.S. in Electronics Engineering from Hanyang University in 1987, and M.S. and Ph.D. degrees in Electrical and Computer Engineering from University of Missouri-Columbia in 1991 and 1995, respectively. He worked with Electronics, and Telecommunication Research Institute (ETRI) from 1995 to 1996, and Korea Telecomm(KT) from 1996 to 1999. He is currently a professor in Chonnam National University in Korea, and the director of Korea Bio-IT Foundry Center at Gwangju. His major research interest is the computational intelligence for image analysis, pattern recognition, network and communication security, bio and medical data analysis.