www.istatistikciler.org
İstatistikçiler Dergisi 5 (2012) 51-64

İstatistikçiler Dergisi

Bayesian Networks and Association Analysis in Knowledge Discovery Process

Derya Ersel

Hacettepe University
Department of Statistics
06800-Beytepe, Ankara, Türkiye
dtektas@hacettepe.edu.tr

Süleyman Günay

Hacettepe University
Department of Statistics
06800-Beytepe, Ankara, Türkiye
sgunay@hacettepe.edu.tr

Abstract

Data mining is a statistical process to extract useful information, unknown patterns and interesting relationships in large databases. In this process, many statistical methods are used. Two of these methods are Bayesian networks and association analysis. Bayesian networks are probabilistic graphical models that encode relationships among a set of random variables in a database. Since they have both causal and probabilistic aspects, data information and expert knowledge can easily be combined by them. Bayesian networks can also represent knowledge about uncertain domain and make strong inferences. Association analysis is a useful technique to detect hidden associations and rules in large databases, and it extracts previously unknown and surprising patterns from already known information. A drawback of association analysis is that many patterns are generated even if the data set is very small. Hence, suitable interestingness measures must be performed to eliminate uninteresting patterns.

Bayesian networks and association analysis can be used together in knowledge discovery. As association rules are used to create Bayesian networks, interestingness measures to determine interesting patterns can be established by Bayesian networks. In this study, this mutual utilization between Bayesian Networks and association analysis is explained and an illustration over a real life problem is presented.

Keywords: Bayesian networks, association analysis, interestingness measures, frequent itemsets.

Özet

Bilgi Keşfi Sürecinde Bayesci Ağlar ve Birliktelik Analizi

Veri madenciliği, büyük veri kümelerinden yararlı bilginin, bilinmeyen örüntülerin ve ilginç ilişkilerin ortaya çıkartıldığı istatistiksel bir süreçtir. Bu süreçte, pek çok istatistiksel yöntem kullanılabilir. Bu yöntemlerden ikisi Bayesci ağlar ve birliktelik analizidir. Bayesci ağlar, bir veri tabanında yer alan raslantı değişkenlerinin bir kümesindeki olasılıksal ilişkileri kodlayan grafiksel modellerdir. Hem nedensel hem de olasılıksal özelliklere sahip olduğundan Bayesci ağlar ile veri ve uzman bilgisi kolaylıkla birleştirilebilir. Bayesci ağlar ayrıca, ilgilenilen problemin kesin olmayan tanım kümesi hakkındaki bilgiyi temsil etmek için kullanılır ve güçlü çıkarsamaların yapılmasını sağlar. Birliktelik analizi, büyük veri tabanlarındaki gizli birlikteliklerin, yararlı kuralların ve şaşırtıcı örüntülerin ortaya çıkartılmasını sağlayan bir yöntemdir.Birliktelik analizinin bir kusuru, veri kümesi çok küçük olsa dahi çok sayıda örüntünün ortaya çıkartılmasıdır.Bu nedenle, bu örüntülerden ilginç olmayanların elenmesi için ilginçlik ölçümleri kullanılmalıdır.

Bayesci ağlar ve birliktelik analizi, bilgi keşfi sürecinde birlikte kullanılabilir.Birliktelik kuralları, Bayesci ağların oluşturulmasında kullanılırken, ilginç örüntülerin belirlenmesinde kullanılan ilginçlik ölçümleri de Bayesci ağlar yardımıyla oluşturulabilir.Bu çalışmada, Bayesci ağlar ve birliktelik analizinin bu karşılıklı kullanımı sunulmuş ve bir veri kümesi üzerinden elde edilen sonuçlar tartışılmıştır.

Anahtar sözcükler: *bayesci ağlar, birliktelik analizi, ilginçlik ölçümleri, sık gözlenen nesne kümeler.*

1. Introduction

In recent years, digital data access and storage techniques have developed rapidly and this has caused very large databases to emerge. So, the occurrence of large databases has created the need for information retrieval [6]. Data mining, in the widest sense, can be defined as extracting useful information from large databases and it is an interdisciplinary field that includes statistics, machine learning, pattern recognition, visualization techniques [10]. Several statistical methods are used to determine the relationships between variables in data mining. In this study, two of these methods - Bayesian networks and association analysis - are discussed.

Bayesian networks, introduced in the early 1990s, are Directed Acyclic Graphs (DAGs) that encode probabilistic relationships between random variables. They provide to model joint probability distribution of a set of random variables efficiently and to make some computations from this model. Lately, Bayesian network shave emerged as an important method which adds uncertain expert knowledge to the system [1, 6]. Generally, two cases are dealt with when analyzing Bayesian networks: inference and learning. In this study, learning problem of Bayesian networks is interested. In literature, many learning algorithms have been proposed. As Bayesian networks can be learned directly from data, they can also be learned from expert opinion. Besides, the results obtained from association analysis can be used to learn and update Bayesian networks [8].

Association analysis is one of the descriptive models used in data mining. It was introduced in 1993 by Agrawal and Friends [3]. The aim of this analysis is to examine items that are seen together frequently in data set and to reveal patterns that help decision making. These patterns are represented as "association rules" or "frequent itemsets" in association analysis. There are two important points to be considered in association analysis. The first point is that obtaining the patterns is complicated and time consuming when data set is large. The second one is that patterns found by association analysis can be deceptive since some relationships may arise by chance. Association algorithms should be constituted by taking into account these situations [10, 15]. A problem encountered in association analysis is that a great number of patterns are generated even if data set is small, so millions of patterns can be obtained when data set is large. To figure out this problem, patterns obtained by association analysis should be evaluated according to their interestingness levels and the patterns which are found uninteresting according to these measures should be eliminated. Interestingness levels are measured by interestingness measures. These measures are categorized into "objective interestingness measures" and "subjective interestingness measures". While objective measures are based on data and structure of pattern, subjective measures are also based on expert knowledge in addition to data and structure of the pattern [15]. Subjective interestingness measures are generally specified through belief systems. Since Bayesian networks are belief systems, these measures can be specified over them [8].

In this study, usage of Bayesian networks to generate a subjective interestingness measures and usage of association analysis to learn Bayesian networks are explained. To present this mutual utilization, in the second section association analysis is presented and interestingness measures are introduced. In the third section Bayesian networks are described briefly. Learning Bayesian networks from interesting patterns and creating interestingness measures from Bayesian networks are explained in the fourth section. Finally, an illustration over 1987 National Indonesia Contraceptive Prevalence Survey is presented.

2. Association analysis

The aim of association analysis is discovering interesting patterns hidden in a database. These patterns are indicated either as "association rules" or "frequent itemsets". Let $I=\{i_1,\ldots,i_d\}$ is the set of all items and $T=\{t_1,\ldots,t_N\}$ is the set of all transactions in the database. Each transaction t_i is a subset of items from I. An itemset is a set of one or more items from I [3].

Let X and Y are disjoint itemsets. An association rule is an expression of the form $X \rightarrow Y$ or $Y \rightarrow X$. Strength of a rule is measured by "support" and "confidence". For the rule $X \rightarrow Y$, support determines the frequency of the items in $X \cup Y$ in a given data set and confidence determines how often items in Y appear in the transactions that contain the items of X. These two measures are defined mathematically as follows.

Support: $s(X \to Y) = \dfrac{\sigma(X \cup Y)}{N}$ (1)

Confidence: $c(X \to Y) = \dfrac{\sigma(X \cup Y)}{\sigma(X)}$ (2)

where N is the total number of transactions[3, 15]. Association analysis determines the rules whose supports and confidences are bigger than "minimum support (minsup)" and "minimum confidence (minconf)" values respectively. The common strategy applied in association analysis algorithms can be given with two stages:

1. *Generating frequent itemsets:* Itemsets with bigger supports than minsup are obtained.

2. *Generating rules:* From among the frequent itemsets obtained at the first stage, the rules with bigger confidences than minconf are determined [10, 15].

The widely used algorithm to obtain association rules is "Apriori" algorithm. This algorithm depends on A priori principal. According to this principal, if an itemset is frequent, all subsets of this itemsets are also frequent. Similarly, if an itemset is infrequent, all supersets of this itemsets are also infrequent. This algorithm prevents unnecessary search of itemsets and hence improves the efficiency.

After determination of the patterns with two steps above, data sets should be re-examined to eliminate uninteresting patterns. To specify interesting patterns, objective or subjective interestingness measures are used. Although objective measures which depend on data and structure of the patterns are useful to determine interesting patterns, they don't contain all information about pattern discovery process. Thus, subjective measures which depend also on expert knowledge are needed to determine interestingness of patterns.

Objective interestingness measures are based on the statistics obtained from the data. Patterns that contain sets of mutually independent items or few transactions are not interesting according to these measures because these relationships generally indicate pseudo relationships in the data. Some objective interestingness measures are listed in Table 1 [5, 14, 15].

Table 1. Some objective interestingness measures.

	Measure	**Equation**
1	Support	$P(X,Y)$
2	Confidence	$\max\left(P(Y \mid X), P(X \mid Y)\right)$
3	ϕ coefficient	$\left[P(X,Y) - P(X)P(Y)\right] \big/ \sqrt{P(X)P(Y) - (1-P(X))(1-P(Y))}$
4	Odds ratio	$\left[P(X,Y)P(\bar{X},\bar{Y})\right] \big/ \left[P(X,\bar{Y})P(\bar{X},Y)\right]$
5	Yule's Q	$\dfrac{P(X,Y)P(\overline{XY}) - P(X,\bar{Y})P(\bar{X},Y)}{P(X,Y)P(\overline{XY}) + P(X,\bar{Y})P(\bar{X},Y)}$
6	Interest	$P(X,Y) \big/ \left[P(X)P(Y)\right]$

The second type of interestingness measures used in association analysis is subjective. Subjective interestingness measures are created with respect to expert knowledge. So, a pattern which is interesting for an expert may not be interesting for another expert. If a pattern is neither unexpected nor actionable, it is not subjectively interesting. Unexpectedness is about beliefs. It can be intuitively said that the more a pattern contradicts with the belief system, the more unexpected (so, more interesting) the pattern is. Therefore, belief system must be described before interestingness of patterns are determined. The interestingness of a pattern "f" related to belief system "S" can be generally defined as,

$$I(f,S)=\sum_i \frac{\left|P(b_i|f,\phi)-P(b_i|\phi)\right|}{P(b_i|\phi)} \tag{3}$$

where each b_i indicates a belief. This measure of interestingness evaluates how much the pattern "f" changes the belief level [13].

3. Bayesian networks

A Bayesian network (BN) is a DAG that represents joint probability distribution of a set of random variables $V=\{X_1,\ldots,X_n\}$. The network consists of two components. The first component G is a DAG where each node corresponds to a random variable X_i $(i=1,\ldots,n)$ and links between the nodes represent direct dependencies between random variables. This component includes conditional independency assumption that each X_i variable is independent from its non-descendants given its parent variables. The second component Θ is set of all parameters in the network. A parameter in the network is defined as $\Theta_{x_i|\pi_i}=P_{BN}(X_i|\pi_i)$ for each x_i state of X_i given the parent set π_i. Hence, a BN for the variable set V defines a joint probability distribution for V and this distribution is obtained by using the Eq (4). Obtaining joint probability distribution by this way is called as "chain rule" [2, 8].

$$P_{BN}(X_1,\ldots,X_n)=\prod_{i=1}^{n} P_{BN}(X_i|\pi_i)=\prod_{i=1}^{n}\Theta_{X_i|\pi_i} \tag{4}$$

The conditional probability distribution $P_{BN}(X_i|\pi_i)$ for X_i is defined as local probability distribution for X_i .If X_i does not have any parent in G, its local probability distribution corresponds to its marginal probability distribution $P_{BN}(X_i)$. Small sets of parents are always preferred in Bayesian network models [2, 5].

To make inferences from Bayesian networks, marginal probability distributions of variables and variable sets are needed to be obtained. One of the most frequently used methods to find marginal distributions is bucket elimination. This method is efficient and easy to apply [2, 12]. Bucket elimination depends on the distributive law. In this method firstly, proper variable ordering is specified and then, the distributive law is implemented repeatedly to simplify the summation [8].

As an illustration, let the joint probability distribution of the variable set $K=\{X,Y,Z\}$ be written as in Eq (2),

$$P(X,Y,Z)=P(X)P(Y|X)P(Z|X) \tag{5}$$

and let marginal probability distribution of X be needed to be found. In this case, the summation below must be calculated.

$$\begin{aligned} P(X) &= \sum_{Y}\sum_{Z} P(X)P(Y|X)P(Z|X) \\ &= P(X)\left[\sum_{Y} P(Y|X)\right]\left[\sum_{Z} P(Z|X)\right] \end{aligned} \tag{6}$$

Each single probability or each summation of a variable over the product of its child buckets is interpreted as a bucket in Eq (6). In this method, determining right variable ordering is an important problem [8].

4. Mutual utilization between Bayesian networks and association analysis

As mentioned before, Bayesian networks and association analysis can be used together in knowledge discovery. While Bayesian networks are used to generate a subjective measure, interesting patterns obtained via association analysis are used to learn Bayesian networks.

4.1 Using Bayesian networks to generate a subjective interestingness measure

Subjective interestingness measures are created with respect to belief systems. Since Bayesian networks are belief systems and they can be used to create a subjective measure. So, the most different patterns from the past information (belief) indicated by a Bayesian network are considered interesting. There are some methods suggested in the literature to determine interesting patterns by using Bayesian networks. Two of these methods were suggested by Jaroszewicz and Simovici [8] and Malhas and Aghbari [11].

Jaroszewicz and Simovici [8] define interestingness of an itemset as absolute difference between its supports estimated from data and Bayesian network. If this difference for an itemset is bigger than a given threshold, this itemset is considered interesting. In this method, interesting itemsets are determined instead of association rules. Direction of rules is specified according to user's experience [5].

Let BN be a Bayesian network over the variable set $V = \{X_1, \ldots, X_n\}$, Q be a subset of V $(Q \subset V)$ and (Q,q) be an itemset $(q \in \mathrm{Dom}(Q))$. The support of the itemset (Q,q) estimated from the data is,

$$\mathrm{support}(Q,q) = P_Q(q) \tag{7}$$

To compute Bayesian support of this itemset, firstly, the joint probability distribution of V must be calculated from Eq.(4). Then, probability distribution of $Q \subset V$ must be calculated using bucket elimination method $\left(P_Q^{BN} = \left[P_V^{BN}\right]^{\downarrow Q}\right)$. So, the support of itemset (Q,q) estimated from BN is,

$$\mathrm{support}_{BN}(Q,q) = P_Q^{BN}(q) \tag{8}$$

As a result, the interestingness of (Q,q) according to BN is calculated with the equation below.

$$I(Q,q) = \left|P_Q(q) - P_Q^{BN}(q)\right| \tag{9}$$

An itemset whose interestingness is bigger than a user specified threshold (w) is w-interesting.

Since dependencies in Bayesian networks are modelled over variables not itemsets, the interestingness of variable sets must be identified. Jaroszewicz and Simovici [8] defined it as,

$$I(Q) = \max_{q \in \mathrm{Dom}(Q)} I(Q,q) \tag{10}$$

Similarly, if $I(Q) \geq w$, then Q is w-interesting.

Another subjective interestingness measure generated using Bayesian networks were suggested by Malhas and Aghbari [11]. This measure is the sensitivity of the Bayesian network to the patterns discovered and it is obtained by assessing the uncertainty-increasing potential of a pattern on the beliefs of Bayesian network. The patterns having the highest sensitivity value is considered the most interesting patterns. In this approach, mutual information is a measure of uncertainty. Sensitivity of a pattern is the sum of the mutual information increases when a pattern enters as an evidence/finding to the Bayesian network.

Mutual information is an indicator of how much information can be obtained about a random variable by observing another random variable. When the variable Y is instantiated to y, the residual uncertainty regarding the true value of a target variable X is

$$H(X|y) = -\sum_{x} P(x,y)\log P(x,y) \tag{11}$$

The mean residual uncertainty in X over all possible values of Y can be calculated as

$$H(X|Y) = \sum P(x,y)\log\frac{P(y)}{P(x,y)} \tag{12}$$

As a result, total uncertainty-reducing potential of Y (mutual information) is

$$M(X,Y) = H(X) - H(X|Y) \tag{13}$$

Let BN be a Bayesian network over the variable set V, $Q \subset V$, $q \in \mathrm{Dom}(Q)$ and (Q,q) be an itemset. Also let R be a query node and F be a finding node. In this case, the sensitivity (S) of the itemset (Q,q) with respect to BN can be calculated as

$$S(Q,q) = \sum_{m=1}^{N}\sum_{n=1}^{N} M_{new}(R_m;F_n) = \sum_{m=1}^{N}\sum_{n=1}^{N} H_{new}(R_m) - H_{new}(R_m|F_n) \tag{14}$$

The sensitivity of the variable set Q is given by

$$S(Q) = \max_{q \in \mathrm{Dom}(Q)} S(Q,q) \tag{15}$$

Variable sets whose sensitivities are less than a user specified threshold (w) can be pruned [11].

4.2 Using interesting patterns to learn Bayesian networks

Learning structure and parameters in Bayesian networks is an important problem in literature. Expert knowledge is generally used to solve this problem. However, it is not always possible to reach the appropriate expert opinion. In current applications, data set is exploited to learn Bayesian network because of the lack of expert opinion. Learning problem in Bayesian networks are separated into "parameter learning" and "structure learning". In this section, "structure learning" problem is dealt with [2, 9].

If there is no expert knowledge about the structure of Bayesian network, the data is used to learn the DAG structure that best describes the data. In principle, in order to find the best DAG structure for the variable set V, all possible DAG representations for V should be established and compared. However, it is hard to create all DAGs for V since number of DAGs increases exponentially in number of variables. So, various approaches to find the best DAG suitable a variable set have been proposed [2].

An approach to determine the DAG structure is to make use of interesting patterns obtained from association analysis over the same data set. This method can be applied in two different ways. If a DAG structure is not created before, variables and links which should be included in DAG are decided according to interesting patterns. If a DAG structure is created before, it can be updated by adding/deleting links and/or adding adding/deleting variables [5, 8]. In this approach, the interestingness measure to be used is firstly determined according to structure of the data, structure of the pattern and presence of expert knowledge. Patterns which have higher interestingness values than a given threshold are specified as interesting patterns. Next, interesting patterns which are mostly duplicated and/or which contain interested variables are detected. According to these patterns, variables which should be included in DAG structure and links between these variables are decided. If a DAG structure exists, the variables duplicated less and not connected with interested variables are eliminated [4].

5. Illustration

In this section, we analyse a dataset which is subset of the 1987 National Indonesia Contraceptive Prevalence Survey. This data set is obtained from UCI Machine Learning Repository [1]. Target population of this survey is married women who are either not pregnant or who do not know if they are at the time of interview.

In the first stage of this section, our problem is to predict womens' current contraceptive method choice (no use, long-term methods, short-term methods) based on their demographic and socio-economic characteristics. Variables are introduced in Table 2 [1]. In original data, "wife's education, husband's education, wife's religion, wife is now working, husband's occupation, standart of living index, media exposure, contraceptive method used" are categorical variables and they remain same in Table 2. However, "wife's age, number of children ever born" is continuous variables in original data and we categorize them in order to use in association analysis and Bayesian networks. Each of these variables is categorized into three levels and they are shown in Table 2.

Table 2. Some variables in 1987 National Indonesia Contraceptive Prevalence Survey.

Variables	Levels
Wife's age	[16, 26] ; [27, 37] ; [38, 49]
Wife's education	Low , Med-, Med+, High
Husband's education	Low , Med-, Med+, High
Number of children ever born	No children ; 1 to 3 children ; 4 or more children
Wife's religion	Non-islam ; Islam
Wife is now working	Yes ; No
Husband's occupation	Low , Med-, Med+, High
Standart of living index	Low , Med-, Med+, High
Media exposure	Good ; Not good
Contraceptive method used	No-use ; Long-term ; Short-term

To predict current contraceptive method, a Bayesian network should be created. To learn Bayesian network for this data set, interesting patterns resulted from association analysis is used. To do this, firstly we create a Bayesian network for these variables according to our (non-expert) prior knowledge (Figure 1) and then, update this network according to interesting patterns. In this study, we create the first Bayesian network with all variables and then update is realized by deleting variables, adding/deleting links and changing the direction of existing links.

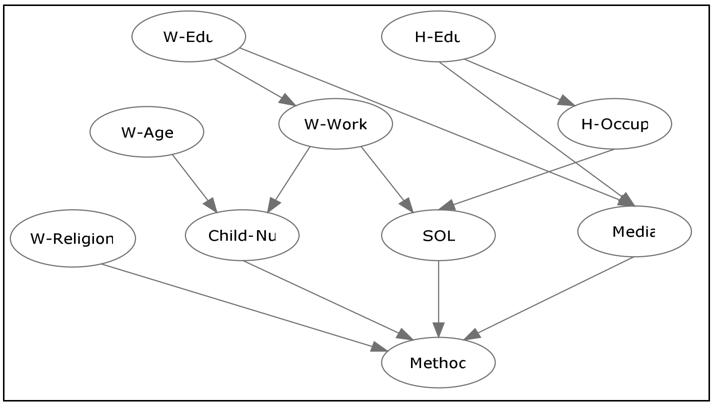

Figure 1. Bayesian network for all variables.

The Bayesian network in Figure 1 can include unnecessary nodes and links to predict the contraceptive method. So, unnecessary and time-consuming computations are performed. Here, association analysis is carried on to reduce number of nodes and links, hence to reduce computations. In this study, to prune this Bayesian network, support and confidence measures are used. To perform association analysis, we use Clementine 11 data mining software tool and determine the most interesting rules according to support and confidence measures. The results that we are obtained are given in Table 3. Interestingness thresholds for support and confidence measures are taken as 0.20 and 0.90 respectively. Since the problem is to predict the current contraceptive method used and the most interesting rules do not contain this variable, the most interesting rules contain "contraceptive method used" are examined separately. To find the most interesting rules containing "contraceptive method used" variable, we perform association analysis again and take the thresholds for confidence and support as 0.85 and 0.10 respectively. Hence, we find the most interesting rules containing the variable "contraceptive method used" for new thresholds. The obtained results are shown in Table 4. So, Table 4 provides to investigate relationships "contraceptive method used" with other variables and Table 3 provides to investigate relationships among other variables related to "contraceptive method used".

Table 3. The most interesting rules according to support and confidence measures (confidence ≥ 0.90 , support ≥ 0.20)

Antecedent	Consequent	Confidence	Support
Wife's education=high	Husband's education= high	0.94281	0.36391
Wife's education= high Media exposure=good	Husband's education= high	0.94221	0.36524
Wife's education= high Wife's religion=islam	Husband's education= high	0.93318	0.27495
Wife's education= high Wife's religion=islam Media exposure=good	Husband's education= high	0.93256	0.27223
Husband's occupation=low	Husband's education= high	0.91514	0.27088
Husband's occupation= low Media exposure=good	Husband's education= high	0.91726	0.26341
Wife's education= high Wife's now working=no	Husband's education= high	0.94349	0.26069
Wife's education= high Wife's now working=no Media exposure=good	Husband's education= high	0.94293	0.25798
Wife's education= high Std-of-living index=high	Husband's education= high	0.96891	0.2539
Wife's education= high Std-of-living index=high	Husband's education= high	0.96875	0.25255

Media exposure=good			
Wife's education= high Number of children=[1, 3]	Husband's education= high	0.96143	0.23693
Wife's education= high Number of children=[1, 3] Media exposure=good	Husband's education= high	0.96111	0.23489
Husband's occupation=low Wife's religion=islam	Husband's education= high	0.90385	0.22335
Husband's occupation=low Wife's religion=islam Media exposure=good	Husband's education= high	0.90265	0.21656
Husband's occupation=low Wife's education= high	Husband's education= high	0.98071	0.20706
Husband's occupation=low Wife's education= high Media exposure=good	Husband's education= high	0.98052	0.20502
Wife's education= high Wife's now working=no Wife's religion=islam	Husband's education= high	0.93671	0.20095
Husband's occupation=low Wife's now working=no	Husband's education=high	0.90769	0.20027

Table 4. The most interesting rules including the variable "contraceptive method used" (confidence ≥ 0.85, support ≥ 0.10).

Antecedent	Consequent	Confidence	Support
Contr.method used=nouse Wife's education=high	Husband's education= high	0.96000	0.11405
Contr.method used=nouse Wife's education= high Media exposure=good	Husband's education= high	0.95954	0.11270
Contr.method used=long Wife's education=e4	Husband's education= high	0.95652	0.13442
Contr.method used=long Wife's education= high Media exposure=good	Husband's education= high	0.95652	0.13442
Contr.method used=short Wife's education= high	Husband's education= high	0.91282	0.12084
Contr.method used=short Wife's education= high Media exposure=good	Husband's education= high	0.91099	0.11813
Contr.method used=long Number of children=2	Husband's education= high	0.86047	0.10048
Contr.method used=short Wife's age=1	Wife's now working=no	0.85795	0.10251

According to interesting rules given in Table 3 and Table 4, some nodes and links can be deleted from Figure 1 and new links can be added to this network. For example, "wife's age" variable is seen only once in Table 4 and it is not seen in Table 3, so we remove this variable from the Bayesian network. For similar reasons, we eliminate "number of children ever born", "wife's religion" and "standart of living index" variables. Since rest of the variables appear more frequently in Table 3 and Table 4, we don't remove these variables from the network. In addition, we add new links from "husband's occupation" to "media exposure" and from "wife's now working" to "contraceptive method used" in Bayesian network. Because, they frequently appear together in Table 3 and Table 4. However, although “wife’s education” and “husband’s education” variables are frequently seen together in these tables, we don’t add a link between these variables. The reason of this is that there is no causal relationship between them.

Consequently, we update the Bayesian network in Figure 1 with respect to previous results. The updated network with the rest of the variables and new links is shown in Figure 2.

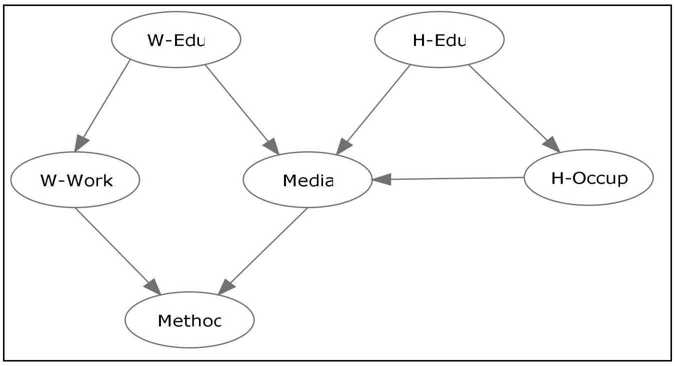

Figure 2. Updated Bayesian network.

After Bayesian network is learned it can be used for inference. In this study, we use NETICA program to calculate some probabilities in this Bayesian network. We found the probabilities for the Bayesian Network in Figure 2 directly from raw data. These probabilities are presented in Figure 3. According to 1987 National Indonesia Contraceptive Prevalence Survey, 42.7% of the women in Indonesia do not use any contraceptive method, 34.7% of them use short-term methods, and 22.6% of them use long-term methods. Similar interpretations can be made from Figure 3.

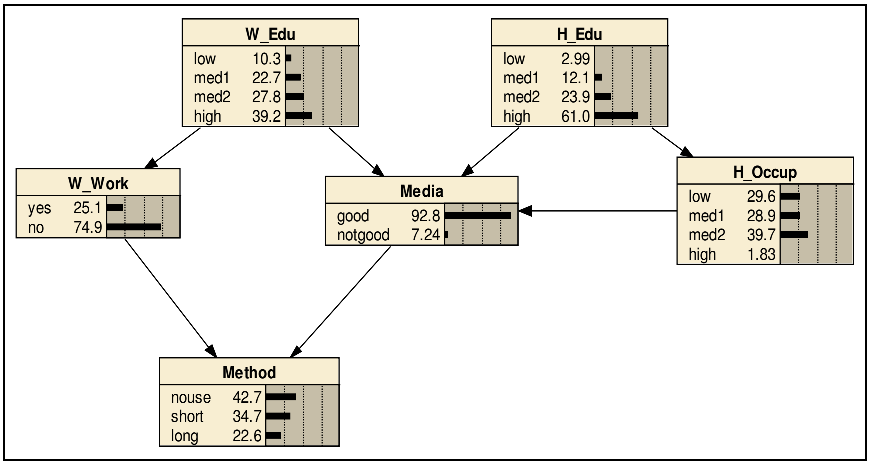

Figure 3. Probabilities for the Bayesian network in Figure 2.

By taking certain values for “method (contraceptive method used)” variable, the probabilities of the other variables can be investigated. We create Figure 3, 4 and 5 by taking the level of method variable “nouse”, “short” and “long” respectively. From these figures it can be said that the probability of “media exposure=good” increases while level of “method” changes from “nouse” to “long”. Similarly, the probabilities of “wife’s education = high” and “husband’s education = high” increase as level of “method” changes from “nouse” to “long”.

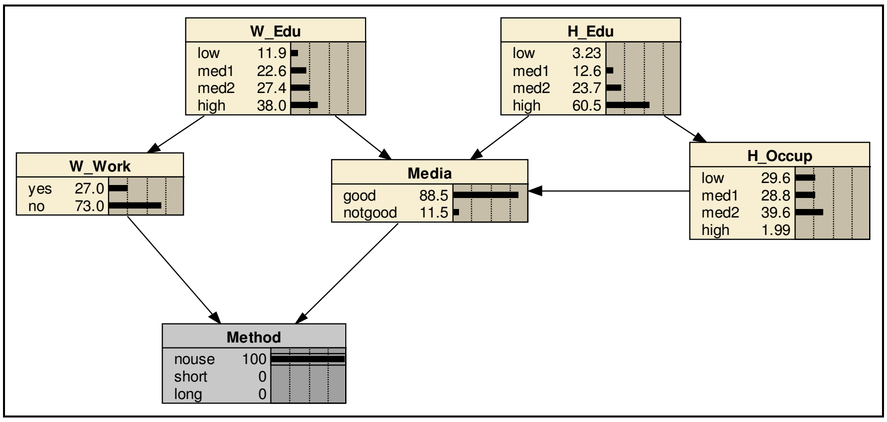

Figure 4. Probabilities for the women who don't use any contraceptive method.

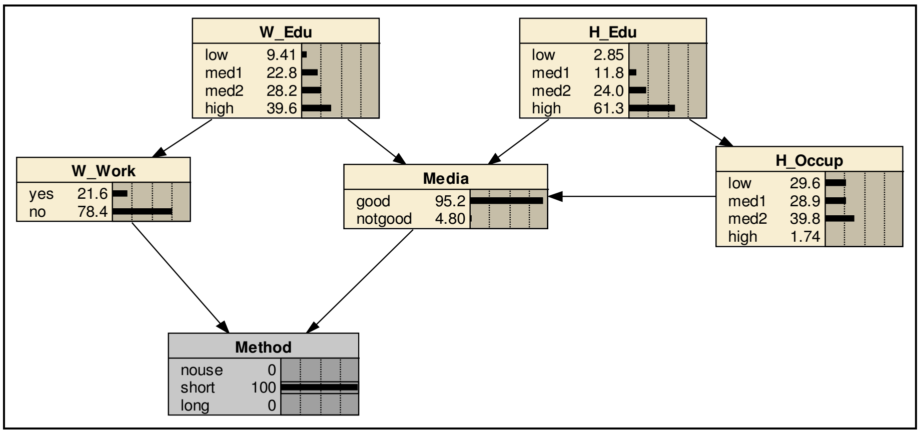

Figure 5. Probabilities for the women who use short-term contraceptive methods.

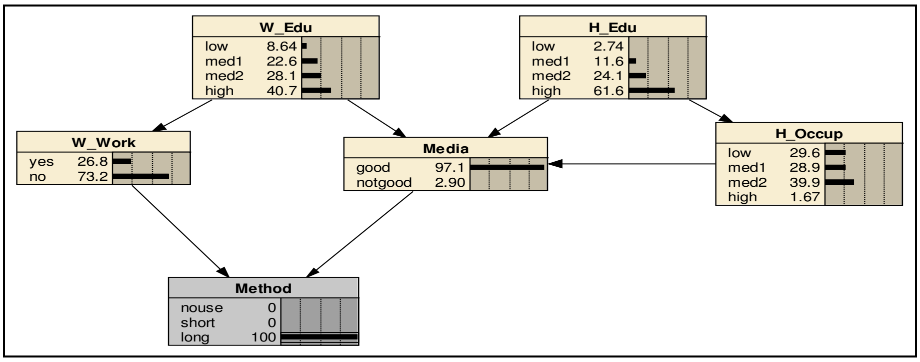

Figure 6. Probabilities for the women who use long-term contraceptive methods.

Conditional probability tables for the variables in the network can also be calculated. We calculate conditional probability table for "method" variable by NETICA (Table 5). It gives the probabilities of levels of "method" variables according to its parent variables.

Table 5. Conditional Probability Table (CPT) for “method”.

Media	W_Work	Method		
		No-use	**Short-term**	**Long-term**
Good	Yes	0.44	0.31	0.25
Good	No	0.40	0.37	0.23
Not good	Yes	0.67	0.18	0.15
Not good	No	0.68	0.25	0.07

The conditional probabilities in Table 5 can be written explicitly as follows:

$$P(\text{Method} = \text{nouse} \mid \text{Media} = \text{good}, \text{W\_Work} = \text{yes}) = 0.44$$
$$P(\text{Method} = \text{short} \mid \text{Media} = \text{good}, \text{W\_Work} = \text{no}) = 0.37$$
$$P(\text{Method} = \text{long} \mid \text{Media} = \text{notgood}, \text{W\_Work} = \text{yes}) = 0.15$$

From Table 5, it can be concluded that “media exposure” has an important effect on “contraceptive method used”. However, “woman’s now working” doesn’t have an important effect on this variable.

In the second stage of this section, the problem is to find interesting patterns using a Bayesian network. Since we create a suitable Bayesian network in the first stage (Figure 2), it can be used to find interesting patterns. Here, we apply Jaroszewicz and Simovici’s[8] approach to find interesting patterns. Python 2.6 program is used to find interesting patterns and their interestingness values (I.V.) according to this method. The results that we are obtained are given in Table 6. In this approach, we take maximum number of variables in an itemset as 5.

Table 6. Interestingness of variables sets according to Bayesian network given in Figure 2.

	VARIABLE SET	**I.V.**		**VARIABLE SET**	**I.V.**
1	[Media,Method]	0.195	28	[W_Edu,H_Edu,W_Work,Media,Method]	0.069
2	[W_Work,Media,Method]	0.194	29	[W_Edu,H_Edu,W_Work,Method]	0.069
3	[W_Work,Method]	0.189	30	[H_Edu,H_Occup,Method]	0.068
4	[H_Edu,Method]	0.143	31	[W_Edu,W_Work,H_Occup]	0.068
5	[H_Edu,W_Work,Method]	0.140	32	[W_Edu,W_Work,Media,H_Occup]	0.068
6	[H_Edu,Media,Method]	0.136	33	[W_Edu,H_Occup,Method]	0.068
7	[H_Edu,W_Work,Media,Method]	0.133	34	[H_Edu,W_Work,H_Occup,Method]	0.064
8	[W_Edu,H_Edu]	0.130	35	[W_Edu,Media,H_Occup,Method]	0.064
9	[W_Edu,H_Edu,Media]	0.128	36	[W_Edu,H_Edu,Media,H_Occup,Method]	0.063
10	[W_Edu,Method]	0.117	37	[W_Edu,H_Edu,H_Occup,Method]	0.063
11	[W_Edu,Media,Method]	0.112	38	[H_Edu,Media,H_Occup,Method]	0.063
12	[W_Edu,W_Work,Method]	0.106	39	[H_Edu,W_Work,Media,H_Occup,Method]	0.059
13	[W_Edu,W_Work,Media,Method]	0.102	40	[W_Edu,W_Work,H_Occup,Method]	0.057
14	[W_Edu,H_Edu,H_Occup]	0.101	41	[W_Edu,W_Work,Media,H_Occup,Method]	0.054
15	[W_Edu,H_Edu,Media,H_Occup]	0.100	42	[W_Edu,H_Edu,W_Work,H_Occup,Method]	0.047
16	[W_Edu,H_Occup]	0.095	43	[Media,H_Occup]	0.019
17	[W_Edu,Media,H_Occup]	0.095	44	[H_Edu,Media]	0.018
18	[W_Edu,H_Edu,Media,Method]	0.092	45	[W_Work,Media,H_Occup]	0.015
19	[W_Edu,H_Edu,Method]	0.092	46	[H_Edu,W_Work,Media]	0.015

20	[W_Edu,H_Edu,W_Work]	0.092	47	[W_Edu,Media]	0.012
21	[W_Edu,H_Edu,W_Work,Media]	0.091	48	[H_Edu,W_Work,Media,H_Occup]	0.011
22	[Media,H_Occup,Method]	0.082	49	[W_Edu,W_Work,Media]	0.010
23	[W_Work,Media,H_Occup,Method]	0.079	50	[W_Work,H_Occup]	0.010
24	[W_Work,H_Occup,Method]	0.077	51	[H_Edu,Media,H_Occup]	0.010
25	[H_Occup,Method]	0.074	52	[H_Edu,W_Work,H_Occup]	0.009
26	[W_Edu,H_Edu,W_Work,H_Occup]	0.072	53	[W_Work,Media]	0.007
27	[W_Edu,H_Edu,W_Work,Media,H_Occup]	0.071	54	[H_Edu,W_Work]	0.002

The most interesting variable set according to Jaroszewicz and Simovici (2004) approach using Bayesian network in Figure 2 is [Media, Method] with interestingness value 0.195. This value gives the absolute difference between this set's support values obtained from data and Bayesian network (Figure 2). This interestingness value is based on the discrepancy between Bayesian network and data. By using different Bayesian networks depending on expert knowledge, more interesting variable sets may be obtained. Bayesian network in Figure 2 can be updated again using the most interesting variable sets given in Table 6, and this updated network can be used again to find interesting patterns. As these steps repeat, the interestingness values of variable sets are reduced. Because, Bayesian network adapts data well and discrepancy between Bayesian network and data decreases [5, 8].

6. Conclusion

Association analysis and Bayesian networks are two methods which are used to accomplish different goals in data mining. Whereas the aim of association analysis is to obtain interesting patterns in a data set, the aim of Bayesian networks is to calculate local probability distributions of the variables by modelling causal relationships between variables. Output of one of these two methods can be used as an input to another method. Interesting patterns determined by association analysis is exploited in learning and updating Bayesian networks. Also, Bayesian networks are exploited to create interestingness measures used in association analysis.

Bayesian networks are generally created according to expert opinion about the problem and the data. Achieving expert opinion is generally difficult and costly. If expert opinion about interested data can not be reached, knowledge obtained from association analysis results can be used to create a Bayesian network. In addition, if expert opinion does not exist but a Bayesian network is created according to non-expert opinion, this Bayesian network can be updated according to association analysis results. Hence, a suitable Bayesian network can be created without the need for expert opinion.

In association analysis, objective interestingness measures are generally used to determine interesting patterns. Interestingness is the incompatibility degree of the pattern to the prior knowledge of the researcher. Objective interestingness measures do not fully comply with this meaning of interestingness. These measures identify patterns frequently seen in data set rather than interesting patterns. However, subjective interestingness measures comply with the meaning of interestingness rather than objective measures. Subjective interestingness measures are defined over expert systems. Bayesian networks are expert systems and they may be used to define a subjective measure. The most different patterns from the knowledge represented by Bayesian networks are specified as the most interesting patterns.

In this study, we explain this mutual utilization between two data mining techniques and perform an illustration to see affects of these two techniques on each other. Illustration section consists of two parts. In the first part, we create a Bayesian network about given data based on our (non-expert) knowledge and then we update this network according to interesting patterns determined by objective interestingness measures. At this stage, subjective interestingness measures can also be used. In the second part of the illustration section, we obtain interesting patterns via a subjective interestingness measure based on Bayesian network created in the first part. Here, a cyclical process can be created. That is to say, interesting patterns obtained can be used to update Bayesian network again, and the new updated network can be used again to determine interestingness measures, and so on.

Using together these two data mining techniques provides to add prior information to knowledge discovery process. Even if this prior information is not received from an expert, suitable results can still be obtained by supporting this non-expert information with data. Hence, there is no need to complicated and time consuming algorithms to learn Bayesian networks and to create interestingness measures, and more suitable results to real world are reached.

References

[1] K. Bache, M. Lichman, 2013, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[2] I. Ben-Gal, 2007, Bayesian Networks,*Encyclopedia of Statistics in Quality &Reliability*,F. Ruggeri, F.Faltin, R. Kenett, R. (eds), Wiley & Sons.

[3] M.W. Berry,M.Browne, 2006,*Lecture Notes in Data Mining*, World Scientific Publishing, Singapore, 222p.

[4] Y. Dong-Peng, L. Jin-Lin, 2008, Research on personal credit evaluation model based on bayesian network and association rules, *Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on* , 457-460.

[5] D. Ersel, 2012,*An Original Combined Interestingness Measure in Association Analysis*, Unpublished PhD Thesis, Hacettepe University Institute of Graduate Studies in Science, Ankara, Türkiye.

[6] D. Hand, H. Mannila, P. Smyth, 2001, *Principles of Data Mining*, The MIT Press, Cambridge, 546p.

[7] D. Heckerman, 1995, Bayesian networks for data mining, *Data Mining and Knowledge Discovery*1, 79-119.

[8] S. Jaroszewicz, D.A. Simovici, 2004, Interestingness of frequent itemsets using Bayesian networks as background knowledge, *Proceedings of the 10th ACM SIGKDD Conference on Knowledge Dicovery and Data Mining*, August 20-25, 2004, New York, USA, 178-186.

[9] F.V. Jensen, 2001, *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York, 268p.

[10] D.T. Larose, 2004, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley Interscience, New York, 222p.

[11] R. Malhas, Z. Aghbari, 2007, Fast discovery of interesting patterns based on Bayesian network background knowledge, *University of Sharjah Journal of Pure and Applied Science*, 4 (3).

[12] K. Murphy, 1998, A brief introduction to graphical models and Bayesian networks, http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html#infer.

[13] A. Siberschatz, A. Tuzhilin, A., 1995, On subjective measures of interestingness in knowledge discovery, *Proceedings of the 1st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 20-21, 1995, Montreal, Canada, 275-281.

[14] P. Tan, V. Kumar, J. Srivastava, 2002, Selecting the right interestingness measure for association patterns, *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* July 23-26, 2002, Edmonton, Alberta, Canada, 32-41.

[15] P.Tan, M. Steinbach, V. Kumar, 2006, *Introduction to Data Mining*, Addison-Wesley, Boston,769p.