Concept Compression in Formal Concept Analysis Using Entropy-Based Attribute Priority

Sumangali K, Aswani Kumar Ch. & Jinhai Li

Taylor & Francis
Taylor & Francis Group

Check for updates

Concept Compression in Formal Concept Analysis Using Entropy-Based Attribute Priority

Sumangali K[a], Aswani Kumar Ch.[a], and Jinhai Li[b]

[a]School of Information Technology and Engineering, VIT University, Vellore, India; [b]Faculty of Science, Kunming University of Science and Technology, Kunming, China

ABSTRACT

Discovering important concepts in formal concept analysis (FCA) is an important issue due to huge number of concepts arising out of complicated contexts. To address this issue, this paper proposes a method for concept compression in FCA, involving many-valued decision context, based on information entropy. The precedence order of attributes is obtained by using entropy theory developed by Shannon. The set of concepts is compressed using the precedence order thus determined. An algorithm namely Entropy based concept compression (ECC) is developed for this purpose. Further, similarity measures between the actual and compressed concepts are examined using the deviance analysis and percentage error calculation on the deviance of input weights of concepts. From the experiments, it is found that the compressed concepts inherit association rules to the maximum extent.

Introduction

The mathematical lattice based framework namely Formal concept analysis (FCA) has emerged as a distinctive tool for knowledge discovery. FCA is a theory of mathematics meant for determining the concepts and their hierarchies that underlie in any information system. For any information system FCA produces a set of concepts and each concept consists of two items, namely extent and intent (Wille 1982). For data analysis, FCA-based approaches are to be essential in the fields of artificial intelligence, decision-making systems, machine learning, knowledge discovery, data mining, expert systems, approximate reasoning and pattern recognition, fault diagnosis, etc., (Priss 2006). FCA requires no prior knowledge or additional information about data unlike probability theory and fuzzy set theory.

In general, any context often produces large number of formal concepts, from which it is very difficult for any user to understand and analyze the context (information system) and such context often contains much redundant

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UAAI.

knowledge. The redundant knowledge present in context wastes the resources and can interfere in the process of decision-making. Hence it is essential to reduce the set of redundant knowledge. To this end, we need to create a simple model of the context. Such models can be created by the structured knowledge of the information system. The structure includes several attributes which may or may not be relevant to the study. Hence it is required to compress the attributes or concepts according to the necessity (Han and Kamber 2006).

There are several methods to produce an attribute ranking by which one can isolate the individual merit of an attribute. Various attribute selection methods which are used in data mining process have been benchmarked by Hall and Holmes (2003). Among the attribute selection techniques we mainly focus on the information gain attribute ranking technique which is simplest and fastest. Entropy is a statistical measure of uncertainty in any information system. The entropy-based attribute selection method is mostly used for the same purpose in which one minimizes the entropy value and thereby maximizes the information gain. The main advantage of this method is that it includes all the attributes in its analysis and is useful when the context is many-valued (MV) and decision-oriented information system. So we deem it is suitable to apply the entropy-based attribute selection method to the problem of current study.

Shannon derived and used the concept of information entropy to determine the level of information gain provided by each attribute of the dataset (Shannon 1948). The notion of entropy plays a vital role in the analysis of uncertain information. Theoretical models of probabilistic distributions fail to depict the scenario of datasets if they contain uncertain information. Hence the notion of entropy has several real-life applications. Rifkin (1980) analyzed the social and political problems by the use of entropy. In (Oruç, Kuruoğlu, and Vupa 2009), from the questionnaire-type dataset the differences in the levels of anxiety between male and female were studied using entropy on few attributes and therein the authors cite various applications of entropy in the field of psychology.

In this article our main focus is on the compression of numerous concepts arising out of a large information system. The motivation for the compression of concepts can be summarized as follows:

- Identification and classification of a set of attributes can determine a set of outcomes that are useful for knowledge reduction process.
- In many information systems the underlying rules and inherent relationships between the attributes are very difficult to handle.
- The extraction of useful knowledge from large number of concepts is a complex process and sometimes such systems do not yield any useful outputs.

Our compression process of concepts is based on the degree of importance of an attribute with respect to other attributes in any context. The degree of importance of an attribute can be determined by the measure of its information gain (Shannon 1948), and hence the attributes can be prioritized accordingly. The primary understanding of the precedence order is to express a type of relative importance of attributes. The precedence order of attributes thus determined plays a vital role in the process of concepts compression. The notion of formal concepts which are compatible with the attribute priorities is introduced in this article. The main benefit of considering such formal concepts is the compression of the set of formal concepts. This leads to a legible structure of formal concepts (clusters) extracted from original concepts. The set of concepts is compressed using the precedence order constraint which is determined out of entropy values. An algorithm namely Entropy based concept compression (ECC) is developed for this purpose and presented in Table 1. Further, similarity measures between the actual and compressed concepts are examined using the deviance analysis and percentage error calculation on the deviance of input weights of concepts. From the experiments, it is found that the compressed concepts subsume association rules to the maximum extent.

The remainder of this article is organized as follows: we emphasize the necessity of knowledge compression and provide an overview of the development of knowledge discovery in back ground study. Also we outline various methodologies adopted with the aim of knowledge compression/reduction

Table 1. Algorithm-entropy based concept compression (ECC).

Input: A decision context $K = (G, M = C \cup \{d\}, I)$
Output: Compressed set of concepts $\mathbf{B}_R(G, M, I)$
　//To compute the entropy of decision attributes
1. $H(D) = -\sum p_k \log_2(p_k)$
　//To compute the entropy of conditional attributes and their information gain
2. for each attribute $C_i \in C$ do $H(C_i) = -\sum_v p_{ij} H(S_{ij})$
Compute $G(C_i) = H(D) - H(C_i)$
　　end for
　//To sort the attributes 'C' in the non-increasing order based on their information gain
3. for each C_i in C do $G(C_i) \geq G(C_{i+1})$//instance with max information gain is ranked 1st in the list
　　rank $G(C_i)$
　　end for
　//To relabel the attributes such that C_i represents the attributes in the ith place of precedence order $C_1 >$
　$C_2 > \dots > C_m$.
4. for each C_i in C do
　　$C_i > C_{i+1}$
　　end for
　//To determine the compressed concepts for consecutive attribute pair
5. for each pair of consecutive attributes C_i, C_{i+1} do
　　$B_{i, i+1}(G, M, I) := \{(A, B) \subseteq B(G, M, I) \mid C_i, C_{i+1} \in B\}, i = 1, 2, \dots, m-1.$
　　end for
　//To obtain the final set of compressed concepts
6. $\mathbf{B}_R(G, M, I) = \overset{m-1}{\underset{i=1}{\cup}} B_{i,i+1}(G, M, I)$

using concepts involved in FCA. The basic notions of FCA and Information Entropy Theory are described in the preliminary section. We next propose Entropy based concept compression (ECC) algorithm to determine the information gain of each attribute and accordingly prioritize the attributes involved in the context. We also mentioned and use certain measurements pertaining to the quality of compressed concepts. Finally, the proposed method of concept compression is illustrated with a real life example context.

Background

Compressing the number of concepts obtained from FCA is an important and challenging task. The process of determining and organizing the set of all formal concepts exhibits exponential behavior in the worst case (Cheung 2004; Kuznetsov 2001). Moreover determining the number of concepts and their hierarchy may become a difficult task in the analysis of the inherent knowledge (Rice and Siff 2001). In specific, the necessary facets which are effectively searched for, can be hidden in innumerable irrelevant details. In many situations such as knowledge extraction, machine learning, ontology construction and others it is vital to have the chance to build an abstracted version of the concepts at a higher level. Indeed, the issue of determining the concepts or concept lattice of a known size and complexity which can reveal the relevant features in the field of application is one of the essential problems of FCA (Dias and Vieira 2013). Some of the contributions toward concept reduction are outlined below.

Kumar and Srinivas (2010) have reduced the size of the concept lattices using fuzzy k–means clustering (FKM). For this purpose, they have reduced the context matrix (object-attribute matrix) and constructed quotient lattices using equivalence relations which are derived by means of FKM Clustering. Ch, Dias, and Vieira (2015) have addressed the issue of knowledge reduction in FCA, based on non-negative matrix factorization of the original context matrix. Li, Li, and He (2014) have attempted to compress a concept lattice arising from incomplete contexts using k-medoids clustering. Recently, Singh and Kumar (2016) reduced a concept lattice using different subset of attributes as information granules. The authors (Belohlavek and Macko 2011; Belohlavek and Vychodil 2009) have proposed a method for concept reduction in which, attribute dependency (AD) formulas are determined from the background knowledge and those concepts which do not obey these AD formulas are removed. Then attributes are assigned weights according to the user preferences. Finally values of formal concepts are obtained; those formal concepts with greater values are regarded more important. Few studies (Li, He, and Zhu 2013; Singh, Cherukuri, and Li 2015) have recently utilized the concept of entropy-based FCA. Singh, Aswani Kumar, and Gani (2016) have concentrated on decreasing the number of formal concepts in FCA with

fuzzy attributes using entropy. Li, He, and Zhu (2013) have proposed entropy-based weighted concept lattice using information entropy and deviance analysis. According to this method weights are determined to the attributes based on the entropy theory, then each concept is identified with an intent average value and deviance analysis is computed for each concept. Further threshold ranges are fixed for deviance measure and the concepts are reduced according to their range.

For reducing a concept lattice's complexity Dias and Vieira (2013) presented the Junction Based Object Similarity approach for replacing groups of similar objects by prototypical ones and introduced two measures such as fidelity and descriptive loss to measure the degree of equivalence between the original formal context and the reduced formal context for evaluating reduction methods. They have illustrated the method from a tuberculosis symptoms database which is carried out at two stages. At the first stage the formal context is reduced by the use of implications expressed in the original formal context. The resulting context is reduced by the use of expert rules (background knowledge) and the effects before and after the reduction are analysed in the next stage. Recently Li et al. (2017) have studied a comparison of reduction in formal decision contexts.

In the study of FCA, reduction of objects and attributes plays an important role in knowledge discovery. Lv, Liu, and Li (2009) proposed a method for attribute reduction from formal concepts using concept lattice. In their analysis, the extents of concepts are studied and formulas concerning significances of attributes are obtained which form the basis for the conclusion if an attribute set is a reduction set of concept lattice. This heuristic information serves the attribute reduction process in FCA. Zhang et al. (2012) have analyzed frequent weighted concept lattices and studied their algebraic properties. Further they have introduced a virtual node into the graphical design of FWCL (frequent weighted concept lattice) in order to retain its completeness. Using the least frequent upper bound and greatest frequent lower bound of the FWCL they further developed an algebraic system of FWCL and studied its properties. They also established its wholeness of knowledge discovery. Another interesting measure towards quality of concept is concept stability. According to Kuznetsov and Obiedkov (2002) in order to evaluate the performance of new algorithm for FCA, a well distinguished set of information sources is required. The stability measure was used by several researchers to reduce the number of formal concepts. Some of them are (Kuznetsov 2007; Roth, Obiedkov, and Kourie 2008). The computation of the stability index is expensive due to the need to calculate the subsets of the extent or intent of each formal concept. Along this line, Babin and Kuznetsov (2012) propose an approximation using Markov chains. The measure of stability of concepts is often used for selecting concepts as biclusters of similar objects.

Analogous to the method of attribute reduction, various methods on object reduction are also studied by researchers. Li et al. (2014) have compressed the objects of a formal decision context using concept lattices to determine the non-redundant decision rules of the original context. Some of the available methods in literature require background knowledge which may not be available and perhaps may not be reliable. Further, the reduction rate of attributes seems to be large in several methods and hence the resulting reduced knowledge may not provide the original information with full entity. In this article, we have developed a new model for concept reduction which requires no prior knowledge and is easy to determine.

Preliminaries

Formal concept analysis

Formal concept analysis is a mathematical framework introduced by Rudolf Wille (1982). It is based on mathematical order theory and on the theory of complete lattices. This model is used for data analysis i.e. for investigating and processing the given information explicitly (Davey and Priestley 2002; Ganter and Wille 1999). FCA has been used in several fields, such as data analysis, knowledge representation and object-oriented class identification tasks (Priss 2006). Any interested reader may refer few references for knowledge representation schemes by use of FCA (Kumar 2011, 2012; Kumar, Ishwarya, and Loo 2015; Kumar and Singh 2014; Kumar and Sumangali 2012; Priss 2006; Shivhare and Cherukuri 2016; Sumangali and Kumar 2013, 2014).

Basic notions of concepts and concept hierarchies

The study of FCA starts with a data collection table called as (*formal*) *context* which is described by a table of crosses. A context K is defined by a triple (G, M, I), where G and M are two sets called objects and attributes respectively, and I is the incidence relation between G and M.

For example in the context shown in Table 3, the elements of G and M represent patients and their symptoms respectively. If an object g in G possesses an attribute m in M it is denoted by $(g, m) \in I$ or gIm and we say that "the patient g has the symptom m".

For $A \subseteq G$ and $B \subseteq M$, we define
$A' := \{ m \in M | (g, m) \in I, \forall g \in A \}$(i.e., the set of attributes common to the objects in A).
$B' := \{ g \in G | (g, m) \in I, \forall m \in B \}$(i.e., the set of objects that have all attributes in B).

Table 2. Many-valued medical diagnosis information system.

Patients	Cough	Vomiting	Cold severing	Nasal bleeding	Temperature	Delirium	Fever
P1	1	2	2	3	5	3	1
P2	2	3	1	3	4	3	2
P3	3	1	2	3	6	3	1
P4	1	2	2	1	4	3	2
P5	3	1	2	1	5	3	2
P6	3	2	1	1	4	3	1
P7	1	1	2	2	4	3	2
P8	2	3	1	2	6	3	1
P9	1	1	3	3	5	3	1
P10	2	2	2	2	5	3	1

Table 3. Formal context derived from Table 2.

	C			V			CS			NB			T			F	T6
	C1	C2	C3	V1	V2	V3	CS1	CS2	CS3	N1	N2	N3	T4	T5	DE	F	T6
P1	X				X			X				X		X		X	X
P2		X				X	X				X	X	X			X	
P3			X	X				X				X			X	X	X
P4	X				X			X		X			X			X	
P5			X	X				X		X				X		X	
P6			X		X		X			X			X			X	X
P7	X				X			X			X		X			X	
P8		X				X	X				X				X	X	X
P9	X			X					X			X		X		X	X
P10		X			X			X			X			X		X	X

In the presented example (in Table 3), *A'* can be thought of as a set of symptoms common to all the patients *A* and, similarly, *B'* is the set of patients possessing the common symptoms in *B*. A concept of the context (G, M, I) is a pair (A, B) where $A \subseteq G$, $B \subseteq M$, such that $B' = A$. We call A the extent and B the intent of the concept (A, B). Thus, a concept is identified by its extent and intent: the extent consists of all objects belonging to the concept while the intent contains all attributes shared by the objects.

For the given context $K = (G, M, I)$, let B(G, M, I) denote the set of all concepts of K. For the concepts (A_1, B_1) and (A_2, B_2) in B(G, M, I) we write $(A_1, B_1) \leq (A_2, B_2)$, iff $A_1 \subseteq A_2$ or $B_1 \supseteq B_2$. The set of formal concepts is organized by the partial order relation $\leq$ (Wille 1982). Hence, the set of all formal concepts of a context K with the *subconcept-superconcept* relation always constitutes a complete lattice and is called as a concept lattice denoted by $L: = (B(K); \leq)$. The infimum and supremum of an arbitrary set of concepts $(O_t, A_t)_{t \in T}$ for any indexed set T in any concept lattice $L: = (B(K); \leq)$ are defined as follows:

$$Inf \wedge_{t \epsilon T}(O_t, A_t) = (\cap_{t \epsilon T} O_t, (\cup_{t \epsilon T} A_t)'')$$

$$Sup \vee_{t \epsilon T}(O_t, A_t) = ((\cup_{t \epsilon T} O_t)'', \cap_{t \epsilon T} A_t)$$

where $O_t = \{A | (A, B) \in (O_t, A_t)\}$ and $A_t = \{B | (A, B) \in (O_t, A_t)\}$.

A concept lattice can be represented graphically by means of lattice diagrams which can be produced by using software tools. ConExp (http://sourceforgenet/projects/conexp) is one of the popular tools for FCA, which has been used in our experiments. Each node of the lattice diagram represents a concept. From a concept lattice, one can also understand the relationship between the concepts. The connection between any two nodes in the concept lattice represents the subconcept-superconcept relation between the corresponding concepts, the upper one being superconcept and the lower one being its subconcept. A hidden concept on which no object/attribute label is shown denotes that neither objects nor attributes are introduced at that node. Instead, it is the combination of objects in the ascending path from the bottom and the attributes in the descending path from the top.

Many-valued contexts and conceptual scaling

Several objects in real life are described by factors or attributes whose values are multiple in nature such as shape, color, country, and gender. Generally, attributes are represented by the presence or absence of some properties. But, this representation scheme is not suitable in case of MV attributes when compared with that of single-valued attributes. To overcome such difficulty in the representation scheme of context, MV context representation scheme is followed. MV contexts can be effectively handled using FCA (Carpineto and Romano 2004; Ganter and Wille 1999). Such contexts usually contain attribute-value in pairs. A MV context consists of quadruple (G, M, V, I), where G is a set of objects, and M is a set of MV attributes, V is a set of attribute values, and a ternary relation I between the sets G, M, and V (i.e., $I \subseteq G \times M \times V$) such that $(g, m, v) \in I$ and $(g, m, w) \in I$ implies $v = w$. The expression $(g, m, v) \in I$ is read as "the attribute m possesses the value v for the object g." The MV attributes can be considered as partial map from G into V, written as $m(g) = v$.

Unlike binary-valued contexts, concept lattices cannot be determined instantly for MV contexts. Hence, one has to convert a MV context into a binary-valued context and this conversion process is termed as *conceptual scaling* according to Ganter and Wille (1989). Such a modified context is known as the derived context. Normally, a conceptual scale is employed on a single attribute m, and in this case the scale forms a basis for the formal context. A scaled MV context $((G, M, V, I), (S_m | m \in M))$, is an ordered pair that consists of MV context (G, M, W, I), and a set of scales $(S_m | m \in M)$. From this MV context, the standard scaling method namely plain scaling creates the derived context,

$K := (G\{(m, n)|m \in M, n \in M_m\}, J)$ where $gJ(m, n) iff m(g) I_m n (g \in G, m \in M, n \in M_m)$.

Several researchers have studied MV contexts using FCA. Messai et al. (2008) have defined MV formal concepts for MV contexts for the first time. In their view MV concept lattices have higher precision levels from which multi-level conceptual clustering can be formed. This approach is useful to improve the efficiency of information retrieval task that possesses complex queries. Assaghir et al. (2009) have presented a method for classifying objects having MV attributes using threshold values on the difference between two attribute values within the same attribute class.

Association rule mining

Besides the schematic lattice representation of knowledge, FCA also produces association rules. Association rules are the simple form of knowledge discovery which are easily understandable and provide information in a compact way. An association rule $P \rightarrow Q$ where $P, Q \subseteq M$ means that those objects having all attributes from P also have all attributes from Q. From any concept lattice the association rules are commonly determined from two bases of implications namely, Duquenne-Guigues (DG) base and Luxenberger base; the former has 100% confidence level while the latter has less than 100% confidence level (Stumme 2002; Zhang and Wu 2011). The DG basis implication has been used in the validation process of the compressed concepts in this article.

Information entropy

In Information theory, entropy denotes the measure of uncertainty present in a random variable. This measure was introduced by Claude E. Shannon (1948) and it denotes the average unpredictability of a random variable which is considered as equivalent to the content of information. It is measured in bits. If "X" is a random variable with set of possible outcomes x_i then $P(x_i)$ is the probability of the outcome $X = x_i$. If "X" is uniformly distributed with n values then $P(X = x_i) = \frac{1}{n}$.

Shannon expressed the information entropy value H as, $H = - \sum_{i=1}^{m} p_i \log_m(p_i)$

In the context of binary variables, measurement is made in bits and hence m is taken as 2.

Some terms and notions in entropy theory

Any decision-based information system can be considered as a formal context $K = (G, M, I)$, where G is the set of objects and $M = \{C \cup D\}$ is the disjoint classes of attributes $C = \{C_1, C_2, ..., C_m\}$ and $D = \{d\}$ called

conditional attributes and decision attributes, respectively. The relation I is called the incidence relation of the context $G \times M$. We next explain the procedure for computing information entropy and information gain.

Let d_k, $k = 1, 2, \ldots, t$. be the decision class of the context. The expected Information or Entropy $H(D)$ in a given context is given by

$$H(D) = -\sum_{k=1}^{t} p_k \log_2(p_k) \tag{1}$$

One can note that log function to the *base2* is used since the information is encoded in bits. For each $k = 1, 2, \ldots, t$, p_k is the probability that an absolute sample belongs to decision class d_k and is equal to s_k/n, where s_k is the number of objects belonging to decision class k and n is the cardinality of the objects belonging to the context.

Let attribute C_i have v distinct values $\{c_{i1}, c_{i2}, \ldots, c_{iv}\}$. Then for each conditional attribute C_i, the set of objects "O" can be partitioned into v subsets S_{ij}, $j = 1, \ldots, v$ such that each set "S_{ij}" consists of those objects having attribute value "c_{ij}". Let "s_{ij}" be the cardinality of the set "S_{ij}" i.e. $sij_= |S_{ij}|$

Let "s_{ijk}" $i = 1, 2, \ldots, m$, be the number of those objects in the set S_{ij} which belong to the decision class d_k. Then the entropy value for each subset "S_{ij}" in the attribute class C_i is given by

$$H(S_{ij}) = -\sum_{k=1}^{t} s_{ijk}/s_{ij} \log_2(s_{ijk}/s_{ij}) \tag{2}$$

Let p_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, v$ denote the probability of an object in Conditional class Ci, to possess the attribute value "cij." Then, $p_{ij} = s_{ij}/n$.

Then expected value of information entropy based on the attribute C_i is defined as follows:

$$H(C_i) = -\sum_{j=1}^{v} p_{ij}H(S_{ij}) \tag{3}$$

The information gain for any attribute C_i, $i = 1, \ldots, m$ is expressed by the formula

$$G(C_i) = H(D) - H(C_i) \tag{4}$$

After computing the information gain of each condition attribute, the attribute with the highest information gain is regarded as the most informative and the most discriminating attribute of the given set. In the proposed method we make use of the information gain values of attributes to rank them. Attributes having higher information gain are considered to be important. Hence we prioritize the attributes in the decreasing order of the values of information gain.

Proposed method—Entropy based concept compression (ECC)

In our concept compression algorithm, we first determine the precedence order for the attribute. Then the method of compression of concepts begins from the original set of concept and then removes those concepts which are not obeying immediate sub-super attribute priority relations that are to be determined by using entropy precedence relation. Finally we get the compressed concept set. The algorithm can be described as follows:

Now, we discuss theoretical aspects of the proposed algorithm. Let "c" be the set of all concepts, "m" be the number of objects, "t" is the number of decision attributes, "n" is the number of conditional attributes, and "v" is the attribute values. To compute the entropy of decision attribute it takes $O(mt)$. Similarly, to compute the entropy of conditional attributes, it takes $O(nvmt)$. As $H(S_{ij})$ takes the complexity $O(mt)$ and calculating P_{ij} requires the complexity $O(v)$. $H(S_{ij})$.. i.e $O(vmt)$. This should be performed for all conditional attributes "n". So, the required total time is $O(nvmt)$. Therefore the entropy of conditional attributes requires $O(nvmt)$ time. By using an efficient sorting algorithm, we can sort the conditional attributes in $O(nlogn)$ time. Relabeling requires $O(n)$ time. Set of compressed concepts can be done in $O(cn)$ time. Finally, we can implement the algorithm in $O(nvmt)+O(cn)$ time.

$\mathbf{B}_R(G, M, I)$ forms the compressed set of concepts based on entropy precedence order. In this algorithm, we determine the information gain and hence the precedence order of the attributes in the first two steps. In the third step we rank and relabel the attributes of the context. The last two steps of the algorithm compress the set of all concepts using precedence order of attributes determined earlier. The next discussion is about the quality of compressed concepts.

Measurement of quality of compressed concepts

General methods of concept reduction/compression similar to those formulated in (Belohlavek and Macko 2011; Cheung 2004; Soldano et al. 2010) have given a view of information often illustrate losses due to the reduction. Hence statistical methods are used in the validation process of the proposed method. In order to achieve this aim we define a suitable similarity evaluation measure of formal concepts based on the average intent weights using the metric of percentage error calculation.

Any context often contains some important attributes. The importance of an attribute is measured in terms of its average intent weight. The statistical fit of any model relative to the mean data can be determined by the use of deviance-based tests. Likelihood ratio tests are based upon the deviance measures and it is used in the process of decision-making. We use these measures to validate the proposed method of concept compression. Li, He,

and Zhu (2013) have suggested intent weight average measure and deviance of an intent which focus on measuring the attributes. For more readability, we provide the definitions of the measures:

Definition 1: Intent weight of a concept

For any object $p(b_i/a_j)$ is called the probability of object a_j to possess the corresponding attribute b_i. The expected (average) value of the attribute b_i is denoted by $E(b_i)$. The probability and the weight value are computed as in (Csiszar 2008).

$$E(b_i) = -\sum_{j=1}^{n} p(b_i/a_j)\log_2 p(b_i/a_j) \tag{5}$$

$$w_i = \frac{E(b_i)}{\sum_{i=1}^{m} E(b_i)} \tag{6}$$

For any concept (A, B), we define $(weight(B))$ as the arithmetic average of the weights of attributes contained in B.

$$weight(B) = \frac{1}{|B|} \sum_{i=1}^{|B|} w_i \tag{7}$$

where $|B|$ is the cardinality of the attribute intent set B. If $|B| = 1$ then B will be a single attribute intent else B is called as multi-attribute intent. Clearly the weight w_i is in a normalized form. We realize that in practical applications $weight(B)$ does not include any significance on the deviation between all the intents. Hence the produced extraction of knowledge may not be conducive to the sense of an interested user. To obviate this shortcoming the following measure is used in the analysis.

Definition 2: Deviance analysis

We introduce an analysis of deviance to determine the significance of the multi-attribute intent weight value. This deviance analysis explores the level of each w_i deviating from $weight(B)$. To this end, we compute the deviation analysis as in (Li, He, and Zhu (2013). The deviation $D(B)$ of the weight values of the attributes in intent B from their mean weight value as defined below is called as the deviance measure of the concept (A, B).

$$D(B) = \begin{cases} 0, & if \ n = 1 \\ \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (w_i - weight(B))^2}, & else \end{cases} \tag{8}$$

Definition 3: Percentage error calculation

The accuracy rate between two models can be studied using percentage error calculation. The variation between the approximate and actual values can be expressed as a percentage of the actual values by using percentage error calculation. It helps us to identify the closeness of the measurement variation between the approximate and actual values of any model. We compute the percentage error of deviance, according to the statistical method used in (Rooker and Gorard 2007). We use the same to compute percentage error between the measures of deviance on compressed and original sets of concepts.

Let d = average deviance value of intents in original concepts and
d' = average deviance value of intents in reduced concepts.
Take $v = \max\{d, d'\}$
The percentage error is defined as

$$\delta = \left| \frac{d - d'}{v} \right| \times 100 \qquad (9)$$

If the absolute error is negligible (less than 10%) then the model is considered to be accurate and the compressed concept is regarded to be more valid (Rooker and Gorard 2007).

Experimental analysis

To demonstrate the proposed method, we consider the standard medical information system used by Tripathy, Acharjya, and Cynthya (2013) shown in Table 2.

The given data are pertinent to the illness of fever. It pertains to 10 patients undergoing medical diagnosis process. Rows of the table are labeled by patients (objects) and columns by symptoms (attributes) that are used to analyze the illness. The $(i, j)^{th}$ entry in the table corresponds to the value of the jth attribute possessed by the ith patient. Attributes of the data table are divided into two disjoint groups called conditional and target attributes. This information system contains six conditional attributes namely *Cough, Vomiting, Cold Severing, Nasal Bleeding, Temperature, Delirium*; and there is a predicted (target) attribute that confirms the illness of fever.

From the classical view of the medical information system, the attribute values which are usually in practice are always, seldom, never, normal, high, and very high. To make the medical diagnosis context simpler, attribute values 1, 2, 3, 4, 5, and 6 are assigned to these attributes, respectively (Tripathy, Acharjya, and Cynthya 2013). Also, the authors have used the values 1 and 2 to predict the presence and absence of fever illness, respectively. However, these assignment of values are optional and do not affect the analysis. Each attribute

varies in its degree of measure with respect to each patient according to the parameter that he/she may earn. In this context the parameter values for the set of attributes are: *Cough* -{*1, 2, 3*},*Vomiting*- {*1, 2, 3*}, *Cold severing*- {*1, 2, 3*}, *Nasal bleeding*- {*1, 2, 3*}, *Temperature* -{*4, 5, 6*}, *Delirium*- {*3*} and finally decision attribute *fever*-{*1, 2*}. We modify the MV medical information system into a binary-valued formal context using conceptual scaling discussed in preliminary section, to obtain a derived context as shown in Table 3.

The abbreviations used are: C—Cough; V—Vomiting; CS—Cold Severing; NB—Nasal Bleeding; T—Temperature; DE—Delirium

In this modified context an attribute having multiple characteristics is split into the number of sub attributes and the attribute values are marked or unmarked by the presence or absence of sub characteristics as the case may be. For instance, the object patient 1 can be seen as described by the six attributes as follows: C—Cough (always); V—Vomiting (seldom); CS—Cold Severing (seldom); NB—Nasal Bleeding (Never); T—Temperature (high); DE—Delirium(never), and F—Fever (yes). The derived context serves as a bit-vector representation, each patient (factor) is assigned a set of 17 bits (one bit for each value of the attributes). We next execute the proposed algorithm described in earlier section for the medical information context presented in Table 3.

Step 1: Information gain of attributes

We first compute the information gain of attributes using the entropy theory to the context presented above. The notations followed below are according to information entropy section.

Here there are two decision classes. Therefore $t = 2$ and $D = \{d_1, d_2\}$,

We observe that $n = 10$, $s_1 = 6$, $s_2 = 4$

Thus, $p_1 = 6/10 = 0.6$, $p_2 = 4/10 = 0.4$

The information entropy of decision attribute fever is computed using *Equation (1)* and we get

$$H(D) = 0.97$$

The label C_i, $i = 1, 2\ldots m$ corresponds to the m attributes of the system. Since we have six such conditional attributes, $m = 6$.

Let us consider the first attribute $C_1 = Cough$ ($i = 1$).

The v distinct values of the attribute C_i are labeled as $\{c_{i1}, c_{i2}, \ldots, c_{iv}\}$. Here C_1 has $v = 3$

distinct attribute values {1, 2, 3}.

With respect to the conditional attribute C_1, the set of objects "O" can be partitioned into $v = 3$ subsets S_{ij}, $j = 1, 2, 3$ according to the attribute values and $sij_= |S_{ij}|$.

Thus $S_{11} = \{p_1, p_4, p_7, p_9\}$, $S_{12} = \{p_2, p_8, p_{10}\}$ *and* $S_{13} = \{p_3, p_5, p_6\}$

Consequently, $s11 = |S_{11}| = 4$, $s12 = |S_{12}| = 3$, $s13 = |S_{13}| = 3$,

We denote by "s_{ijk}" $i = 1, 2, \ldots, m$; $j = 1,2,\ldots, v$; $k = 1,2$ the number of those objects in the set S_{ij} which belong to the decision class d_k.

When $i = 1$, $j = 1, 2, 3$.

We have, $s_{11} = 4$, $s_{111} = 2$, $s_{112} = 2$

$s_{12} = 3$, $s_{121} = 2$, $s_{122} = 1$

$s_{13} = 3$, $s_{131} = 2$, $s_{132} = 1$.

The entropy value for each subset S_i $_j$in the attribute class C_i is computed using Equation (2)

$$H(S_{11}) = -\sum_{k=1}^{2} \frac{s_{11k}}{s_{11}} \log_2 \frac{s_{11k}}{s_{11}}$$

$$= -[0.5 \log_2 0.5 + 0.5 \log_2 0.5] = -1$$

Similarly we get,

$$H(S_{12}) = 0.922$$
$$H(S_{13}) = 0.922$$

Further, the probability of an object in Conditional class $C_{i,}$ to possess the attribute value "c_{ij}" $i = 1,2,\ldots,6$, $j = 1,2, \ldots, v$ is given by $p_{ij} = s_{ij}/n$.

For $i = 1$, $p_{11} = s_{11}/n = 4/10 = 0.4$

$$p_{12} = s_{12}/n = 3/10 = 0.3$$

$$p_{13} = s_{12}/n = 3/10 = 0.3$$

The expected value of information entropy based on the attribute C_i is computed using Equation (3)

Thus,

$$H(C_1) = -\sum_{j=1}^{3} p_{1j}H(S_{1j}) = p_{11}H(S_{11}) + p_{12}H(S_{12}) + p_{13}H(S_{13})$$

$= 0.4 \times 1 + 0.3 \times 0.922 + 0.3 \times 0.922$

$= 0.9532$

We compute the information gain of each condition attribute using Equation (4).

Thus the information gain for the attribute $C_1 = Cough$ is

$$G(C_1) = H(D) - H(C_1)$$

$= 0.97 – 0.9532 = 0.017$.

From this information gain value, we conclude that the conditional attribute $C_1 = Cough$ can reduce the decision-making uncertainty by its gain value "0.017" for classifying information system K. The attribute with the

Table 4. Results obtained during the computation of information gain values.

	C_{cough}	$C_{vomiting}$	$C_{coldsevering}$	$C_{nasalbleeding}$	$C_{temperature}$	$C_{delirium}$
H_1	1	1	0.922	0.922	–	–
H_2	0.922	0.8112	1	0.922	–	–
H_3	0.922	1	0	0.8112	–	0.969
H_4	–	–	–	–	0.8112	–
H_5	–	–	–	–	0.8112	–
H_6	–	–	–	–	0	–
$H(C_i)$	0.9532	0.924	0.8766	0.8777	0.648	0.969
$G(C_i) = H(D) - H(C_i)$	0.017	0.046	0.094	0.093	0.322	0.001

highest information gain is the most informative and the most discriminating attribute of the given set.

In the same manner, the function of other conditional attributes for reducing decision-making uncertainty can be summed up as shown in Table 4.

$H(C_i)$ in the table is the information entropy of the ith conditional attribute involved in the medical diagnosis process. The values of information gain $G(C_i)$ in the last row $H(D) - H(C_i)$ indicate the precision level of an attribute.

Step 2: Entropy-based precedence relation of conditional attributes in the context

An attribute with higher precision value plays a vital role in the decision-making environment and hence ranked high. According to this rule we get, *Temperature > Cold Severing > Nasal Bleeding > Vomiting > Cough > Delirium*. Where *P >Q* implies, *P* is more important than *Q*. The information gain values of attributes presented in Table 4 lead us to decide that the *temperature* symptom is the most important attribute in the decision-making environment. The precedence order of attributes according to the gain measure is:

$$T > CS > NB > V > C > DE$$

From this precedence order we revise the labeling of attributes. Now we get, *C_1 = Temperature, C_2 = Cold Severing, C_3 = Nasal Bleeding, C_4 = Vomiting, C_5 = Cough, C_6 = Delirium*

Step 3: Underlying concepts in the context

Formal concept analysis explores all the concepts present in the given context from which one can determine the intents and extents of each concept. All such concepts form a concept lattice. The whole conceptual structure consists of concepts and relations. Each concept (*A, B*) can be represented by means of a node with its related extent and intent. If a set of objects *A* is attached to some

concept, then all the objects of A lie in extents of all concepts, reachable by ascending paths in the lattice diagram from this concept to unit concept (top element) of lattice. If a set of attributes B is attached to some concept, then the attributes of B occur in intents of all concepts, reachable by descending paths from this concept to zero concept (bottom element) of lattice. The concepts of the context Table 3 are determined using Next closure algorithm (Carpineto and Romano 2004) and listed in Table 5.

Table 5. Concepts derived from the formal context Table 3.

Concept #	({Extent};{Intent})
C1	({φ};{C1, C2, C3, V1, V2, V3, CS1, CS2, CS3, N1, N2, N3, T4, T5, T6, DE})
C2	({P8};{C2, V3, CS1, N2, T6, DE})
C3	({P10};{C2, V2, CS2, N2, T5, DE})
C4	({P2};{C2,V3,CS1,N3,T4,DE})
C5	({P3};{C3, V1,CS2, N3, T6,DE})
C 6	({P9};{C1,V1,CS3, N3, T5, DE})
C7	({P7};{C1, V1, CS2, N2, T4, DE})
C8	({P1};{C1, V2, CS2, N3, T5, DE})
C9	({P5};{C3, V1, CS2, N1, T5, DE})
C10	({P4};{C1, V2, CS2, N1, T4, DE})
C11	({P6};{C3, V2, CS1, N1, T4, DE})
C12	({P8, P10};{C2, N2, DE})
C13	({P8, P3};{T6, DE})
C14	({P8, P2};{C2, V3, CS1, DE})
C15	({P3, P9};{V1, N3, DE})
C16	({P10, P7};{CS2, N2, DE})
C17	({P9, P7};{C1, V1, DE})
C18	({P9, P1};{C1, N3, T5, DE})
C19	({P3, P1};{CS2, N3, DE})
C20	({P9, P5};{V1, T5, DE})
C21	({P10, P1};{V2, CS2, T5, DE})
C22	({P2, P6};{CS1, T4, DE})
C23	({P3, P5};{C3, V1, CS2, DE})
C24	({P1, P4};{C1, V2, CS2, DE})
C25	({P4, P7};{C1, CS2, T4, DE})
C26	({P4, P5};{CS2, N1, DE})
C27	({P4, P6};{V2, N1, T4, DE})
C28	({P5, P6};{C3, N1, DE})
C29	({P2, P8, P10};{C2, DE})
C30	({P7, P8, P10};{N2, DE})
C31	({P1, P2, P3, P9};{N3,DE})
C32	({P2, P8, P6};{CS1,DE})
C33	({P3, P5, P7};{V1, CS2, DE})
C34	({P1, P5, P10};{CS2, T5, DE})
C35	({P1, P4, P7};{C1, CS2, DE})
C36	({P1, P4, P10};{V2, CS2, DE})
C37	({P3, P5, P6};{C3, DE})
C38	({P2, P7, P4, P6};{T4, DE})
C39	({P4, P5, P6};{N1, DE})
C40	({P3, P5, P7, P9};{V1, DE})
C41	({P1, P4, P7, P9};{C1, DE})
C42	({P1, P5, P9, P10};{T5, DE})
C43	({P3, P5, P7, P1, P4, P10};{CS2, DE})
C44	({P1, P4, P6, P10};{V2,DE})
C45	({P1, P2, P3, P4, P5, P6, P7, P8, P9, P10};{DE})
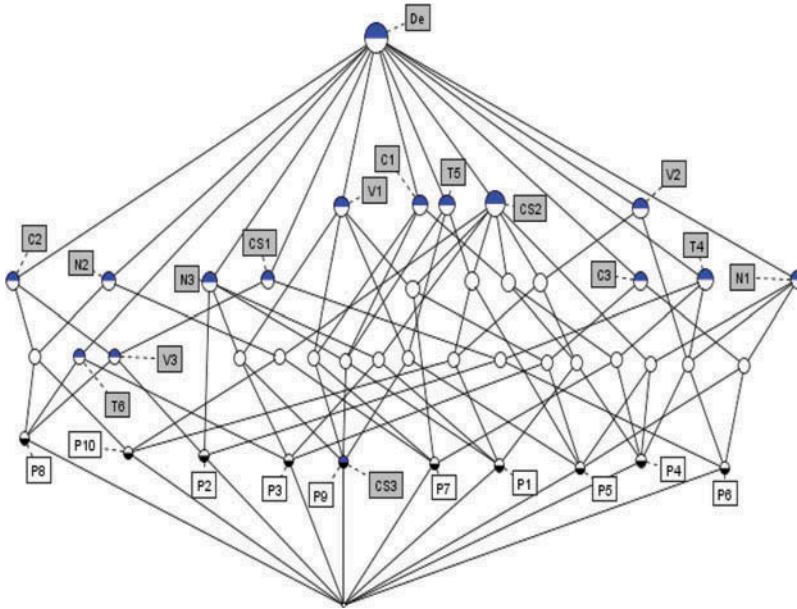
Figure 1. Concept lattice derived from the context Table 3.

These set of all concepts can be represented as a complete lattice under the partial order relation $\leq$ as shown in Figure 1. We can identify from Figure 1 that, the lattice structure is of height 5 and contains 45 concepts connected by 102 edges. Any two connected nodes in the concept lattice represents the sub-superconcept relation between the corresponding concepts, the upper one being the superconcept and the lower one being its subconcept. Attributes disseminate beside the boundaries to the bottom of the lattice while objects disseminate to the top of the lattice.

Step 4: Concepts obeying immediate sub-super attribute priority

For any two consecutive attributes $C_i > C_{i+1}$, We set $\mathbf{B}_{i,\ i+1}(G, M, I) = \{(A, B) \subseteq \mathbf{B}(G, M, I) \,|C_i,\ C_{i+1} \in B\}$, $i = 1, 2, \ldots, m\text{-}1$

In other words the set of concepts $\mathbf{B}_{i,\ i+1}(G, M, I)$ includes the set of concepts from $\mathbf{B}(G, M, I)$ in which both the consecutive attributes C_i, C_{i+1} are present. We list the set of discarded concepts which contain no two consecutive attributes in the above prescribed precedence relation order in Table 6. The precedence relation $T > CS$ implies that attribute T is the immediate predecessor of attribute CS. The compressed concepts for each pairwise precedence relation are given in Table 7.

Table 6. Precedence relation-based discarded concepts.

Immediate predecessors	Removed concepts
C > DE	C13, C15, C16, C 19, C 20, C 21, C 22, C 26, C 27, C 30, C31, C 32, C 33, C34, C36, C38, C39, C40, C42, C43, C44
V > C	C12, C13, C15, C16, C18, C19, C20, C21, C22, C25, C26, C27, C28, C29, C30, C31, C32, C33, C34, C35, C36, C37, C38, C39, C40, C41, C42, C43, C 44
NB > V	C12, C13, C14, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C28, C29, C30, C31, C32, C33, C34, C35, C36, C37, C38, C39, C40, C41, C42, C43, C44
CS > NB	C12, C13, C14, C15, C17, C18, C20, C21, C22, C23, C24, C25, C27, C28, C29, C30, C31, C32, C33, C34, C35, C36, C37, C38, C39, C40, C41, C42, C43, C44
T > CS	C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C26, C27, C28, C29, C30, C31, C32, C33, C35, C36, C37, C38, C39, C40, C41, C42, C43, C44

Table 7. Retained concepts for each precedence relation. $C_{56} = \{(A,B) \in B \, (G, M, I) \,|C, DE \in B\}$.

Immediate predecessors	Retained concepts
C > DE	
$C_{56} = \{(A,B) \in B \, (G, M, I) \,|C, DE \in B\}$	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C14, C17, C18, C23, C24, C25, C28, C29, C35, C37, C41
V > C	
$C_{45} = \{(A,B) \in B \, (G, M, I) \,|V, C \in B\}$	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C14, C17, C23, C24
NB > V	
$C_{34} = \{(A,B) \in B \, (G, M, I) \,| NB, V \in B\}$	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C15, C27
CS > NB	
$C_{23} = \{(A,B) \in B \, (G, M, I) \,| CS, NB \in B\}$	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C16, C19, C26
T > CS	
$C_{12} = \{(A,B) \in B \, (G, M, I) \,|T, CS \in B\}$	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C21, C22, C25, C34

Step 5: Precedence relation-based concept compression

The set of concepts $BR(G, M, I) = \overset{m-1}{\underset{i=1}{\cup}} B_{i,i+1}(G, M, I)$ forms the set of required entropy precedence order-based reduced concepts. It is interesting to note that, the concepts which obey the precedence order relation with respect to at least one of precedence order pairs of attributes are retained in the final reduced set of concepts. In other words those concepts which do not obey the precedence order relation of attributes with respect to any of the precedence order pairs only are removed from the set of concepts.

In this context we obtain 31 concepts using the above formula which constitute the set of reduced concepts. They are: *C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C14, C15, C16, C17, C18, C19, C21, C22, C23, C24, C25, C26, C27, C28, C29, C34, C35, C37, and C41.*

We assume that in any context the attributes present in any object should obey at least any one of the precedence order pair relations. So, the same must hold for the attributes for an intent of any concept. Then the

desired set of final compressed concepts must be the union of such concepts due to each precedence order pair. Thus out of the 45 total number of original concepts, after compression the number of concepts obtained is 31.

Measurement of quality of compressed concepts

In this process we firstly compute measurements pertinent to attributes such as, probability, expectation, and weight of each attribute using Equations (5) and (6). The results are shown in Table 8.

Secondly, we compute the measurements pertinent to concepts, namely intent average value, weight and deviance for each concept using *Equations* (7) and (8). The results are displayed in the Table 9. We tabulate the measurements corresponding to compressed concepts, in Table 10.

Percentage error calculation for validation process

The percentage error of deviance for the compressed concepts from medical diagnosis context is obtained using *Equation* (9), we get

$\sum D(I) = 1.681$ and therefore, $d = 1.681/45 = 0.0373$

and $\sum D(I') = 1.113$, and $d' = 1.113/31 = 0.0359$

where I and I' respectively denote the sets of intents of the original and compressed set of concepts. $v = max\{0.0373, 0.0359\}$

$$\delta = \left| \frac{0.0373 - 0.0359}{0.0373} \right| \times 100 = \frac{0.002}{0.0373} \times 100 = 0.053 \times 100 = 5\%$$

The negligible percentage error explores that those intents which have been removed from the original concepts during the process of concept

Table 8. Acquisition method for the single intent weight value.

Attribute	P(X)	E(X)	W_i
C1	0.4	0.529	0.071
C2	0.3	0.521	0.07
C3	0.3	0.521	0.07
V1	0.4	0.529	0.071
V2	0.4	0.529	0.071
V3	0.2	0.464	0.062
CS1	0.3	0.521	0.07
CS2	0.6	0.442	0.059
CS3	0.1	0.332	0.044
N1	0.3	0.521	0.07
N2	0.3	0.521	0.07
N3	0.4	0.529	0.071
T4	0.4	0.529	0.071
T5	0.4	0.529	0.071
T6	0.2	0.464	0.062
DE	1	0	0

Table 9. The intent weight value and importance deviation value of the original concepts.

Concept #	Intent	Intent average value	Weight (I)	D(I)
C1	C1, C2, C3, V1, V2, V3, CS1, CS2, CS3, N1, N2, N3, T1, T2, T3, DE	0.063	0.063	0.018
C2	C2, V3, CS1, N2, T3, DE	0.056	0.056	0.028
C3	C2, V2, CS2, N2, T2, DE	0.057	0.057	0.028
C4	C2,V3,CS1,N3,T1,DE	0.057	0.057	0.028
C5	C3,V1,CS2, N3, T3,DE	0.056	0.056	0.028
C6	C1,V1,CS3, N3, T2, DE	0.055	0.055	0.029
C7	C1, V1, CS2, N2, T1, DE	0.057	0.057	0.028
C8	C1, V2, CS2, N3, T2, DE	0.057	0.057	0.028
C9	C3, V1, CS2, N1, T2, DE	0.057	0.057	0.028
C10	C1, V2, CS2, N1, T1, DE	0.057	0.057	0.028
C11	C3, V2, CS1, N1, T1, DE	0.059	0.059	0.029
C12	C2, N2, DE	0.047	0.047	0.04
C13	T3, DE	0.031	0.031	0.044
C14	C2, V3, CS1, DE	0.051	0.051	0.034
C15	V1, N3, DE	0.047	0.047	0.041
C16	CS2, N2, DE	0.043	0.043	0.038
C17	C1, V1 DE	0.047	0.047	0.041
C18	C1, N3, T2, DE	0.053	0.053	0.035
C19	CS2, N3, DE	0.043	0.043	0.038
C20	V1, T2, DE	0.047	0.047	0.041
C21	V2, CS2, T2, DE	0.05	0.05	0.034
C22	CS1, T1, DE	0.047	0.047	0.041
C23	C3, V1, CS2, DE	0.05	0.05	0.034
C24	C1, V2, CS2, DE	0.05	0.05	0.034
C25	C1, CS2, T1, DE	0.05	0.05	0.034
C26	CS2, N1, DE	0.043	0.043	0.038
C27	V2, N1, T1, DE	0.053	0.053	0.035
C28	C3, N1, DE	0.047	0.047	0.04
C29	C2, DE	0.035	0.035	0.049
C30	N2, DE	0.035	0.035	0.049
C31	N3,DE	0.036	0.036	0.05
C32	CS1,DE	0.035	0.035	0.049
C33	V1, CS2, DE	0.043	0.043	0.038
C34	CS2, T2, DE	0.043	0.043	0.038
C35	C1, CS2, DE	0.043	0.043	0.038
C36	V2, CS2, DE	0.043	0.043	0.038
C37	C3, DE	0.035	0.035	0.049
C38	T1, DE	0.036	0.036	0.05
C39	N1, DE	0.035	0.035	0.049
C40	V1, DE	0.036	0.036	0.05
C41	C1, DE	0.036	0.036	0.05
C42	T2, DE	0.036	0.036	0.05
C43	CS2, DE	0.03	0.03	0.042
C44	V2, DE	0.036	0.036	0.05
C45	DE	0	0	0

compression have greater deviance measures of intent weights than those of existing ones, which in turn implies that compressed concepts have less deviance of intent weights between them. Therefore we may conclude that the compressed concepts provide the information about the system to the maximum level.

Table 10. The intent weight value and importance deviation value of the retained concepts

Concept #	Intent	Intent average value	Weight(I)	D(I)
C1	C1, C2, C3, V1, V2, V3, CS1, CS2, CS3, N1, N2, N3, T1, T2, T3, DE	0.063	0.063	0.018
C2	C2, V3, CS1, N2, T3, DE	0.056	0.056	0.028
C3	C2, V2, CS2, N2, T2, DE	0.057	0.057	0.028
C4	C2,V3,CS1,N3,T1,DE	0.057	0.057	0.028
C5	C3, V1,CS2, N3, T3,DE	0.056	0.056	0.028
C6	C1,V1,CS3, N3, T2, DE	0.055	0.055	0.029
C7	C1, V1, CS2, N2, T1, DE	0.057	0.057	0.028
C8	C1, V2, CS2, N3, T2, DE	0.057	0.057	0.028
C9	C3, V1, CS2, N1, T2, DE	0.057	0.057	0.028
C10	C1, V2, CS2, N1, T1, DE	0.057	0.057	0.028
C11	C3, V2, CS1, N1, T1, DE	0.059	0.059	0.029
C12	C2, N2, DE	0.047	0.047	0.04
C14	C2, V3, CS1, DE	0.051	0.051	0.034
C15	V1, N3, DE	0.047	0.047	0.041
C16	CS2, N2, DE	0.043	0.043	0.038
C17	C1, V1 DE	0.047	0.047	0.041
C18	C1, N3, T2, DE	0.053	0.053	0.035
C19	CS2, N3, DE	0.043	0.043	0.038
C21	V2, CS2, T2, DE	0.05	0.05	0.034
C22	CS1, T1, DE	0.047	0.047	0.041
C23	C3, V1, CS2, DE	0.05	0.05	0.034
C24	C1, V2, CS2, DE	0.05	0.05	0.034
C25	C1, CS2, T1, DE	0.05	0.05	0.034
C26	CS2, N1, DE	0.043	0.043	0.038
C27	V2, N1, T1, DE	0.053	0.053	0.035
C28	C3, N1, DE	0.047	0.047	0.04
C29	C2, DE	0.035	0.035	0.049
C34	CS2, T2, DE	0.043	0.043	0.038
C35	C1, CS2, DE	0.043	0.043	0.038
C37	C3, DE	0.035	0.035	0.049
C41	C1, DE	0.036	0.036	0.05

One can understand that the proposed method best suits as much as the number of intents for each concept is more than one. In case of concepts having single attribute intents, the proposed method of compression has comparatively little higher rate of percentage error. But, yet the suitability of the proposed method need not be disregarded even on this particular case since the knowledge extracted from this compressed concepts is similar to that of original concepts. As an example one can see that, the compressed concepts also include all the concepts which bound all the association rules. We illustrate this scenario with the help of following example used in (Fan, Liu, and Tzeng 2007).Table 11 is a summary of reviewers' report for ten papers submitted to a journal. The table details ten papers evaluated by means of four attributes: O: originality, P: presentation, TS: technical soundness, and E: overall evaluation (the decision attribute).

The binary-valued formal context derived from the MV context shown in Table 12. Table 13 shows the computation of information gain values. From

Table 11. Many-valued article evaluation information system.

Objects	Originality	Presentation	Technical soundness	Overall evaluation
1	4	4	3	4
2	3	2	3	3
3	4	3	2	3
4	2	2	2	2
5	2	1	2	1
6	3	1	2	1
7	3	2	2	2
8	4	1	2	2
9	3	3	2	3
10	4	3	3	3

Table 12. Formal context derived from Table 11.

	O			P				TS		E			
	O2	O3	O4	P1	P2	P3	P4	T2	T3	1	2	3	4
1			X				X		X				X
2		X			X				X			X	
3			X			X		X				X	
4	X				X			X			X		
5	X			X				X		X			
6		X		X				X		X			
7		X			X			X			X		
8			X	X				X			X		
9		X				X		X				X	
10			X			X			X			X	

Table 13. Results obtained during the computation of information gain values.

	C_O	C_P	C_T
H_1	1	0.91829	—
H_2	1	0.91829	1.55664
H_3	1.5	0	0.91829
H_4	1.5	0	—
$H(C_i)$	1.4	0.55098	1.36514
$G(C_i) = H(D) - H(C_i)$	0.4464	1.29542	0.48126

the computations yield the precedence order of attributes is P > T > O. The formal context given in Table 12 yields the following concept lattice with 27 concepts. These set of all concepts can be represented as a complete lattice under the partial order relation $\leq$ as shown in Figure 2. The compressed and discarded concepts are shown in Table 14.

Further, similar to quality measurements carried out with medical diagnosis context, one can obtain the performance results of the proposed method on article evaluation context. It is found that, in this case, the sum of deviances of the intents of the original and compressed concepts are respectively,

$\sum D(I) = 2.027$ and $d = 2.027/27 = 0.075074$

$\sum D(I') = 1.783$ and $d' = 1.783/18 = 0.099056$

Thus, the percentage error is found to be $\delta \simeq 24\%$.

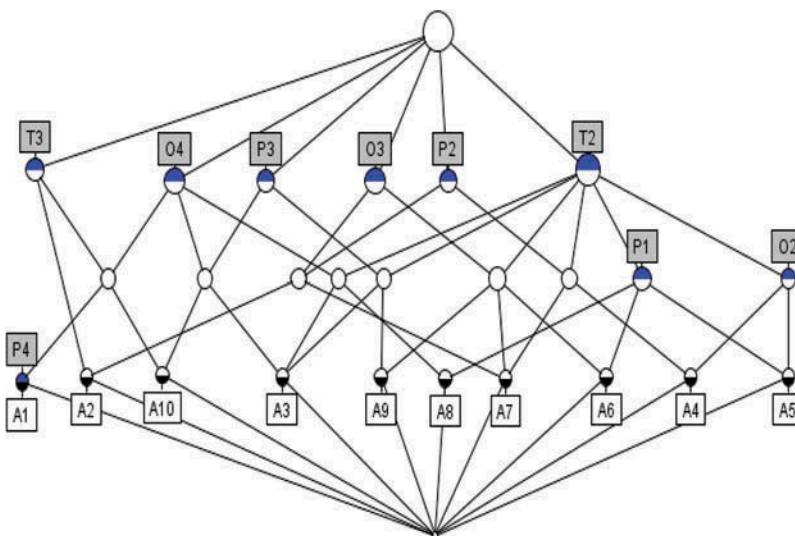**Figure 2.** Concept lattice derived from Table 12..

Table 14. Removed and retained concepts based on precedence relation.

Immediate predecessors	Removed concepts	Retained concepts
T > O	C13, C14, C16, C18, C19, C21, C22, C23, C24, C25, C26, C 27	C1, C2, C3, C4, C5, C 6, C7, C8, C9, C10, C11, C12, C15, C17, C20
P > T	C12, C13, C14, C15, C17, C20, C21, C22, C23, C24, C25, C26, C27	C1, C2, C3, C4, C5, C 6, C7, C8, C9, C10, C11, C16, C18, C19

Underlying association rules after concept compression

We next validate the compressed concepts using association rules (refer preliminary section). For the concept lattice Figure1 obtained from the given context Table 3, there are 53 implications derived according to the DG base. These implications are contained within 18 concepts of FCA. We find that all these 18 concepts are contained in the compressed set of concepts $\mathbf{B}_R(G, M, I)$. Table 15 lists these concepts and their set of implications. The association rules that underlie after the compression of concepts in the case of second example may be considered as an exercise to the readers. After evaluation it is found that there are 17 implications according to the DG base. These implications are contained within 7 concepts of FCA. It is found that all these 7 concepts are contained in the compressed set of concepts $\mathbf{B}_R(G, M, I)$. Therefore one can form a basis for further reduction of concepts.

Conclusions

There are several methods available in the literature for knowledge reduction in FCA. In this article, we concentrate on compressing the concepts of many-valued

Table 15. Concepts and the (subsisting) set of implications derived from Table 3.

Concept #	Implications with 100% confidence
C2	CS1 N2 DE → C2 V3 T6
	C2 T6 DE → V3 CS1 N2
	CS1 T6 DE → C2 V3 N2
	N2 T6 DE → C2 V3 CS1
C3	C2 V2 DE → CS2 N2 T5
	C2 CS2 DE → V2 N2 T5
	V2 N2 DE → C2 CS2 T5
	C2 T5 DE → V2 CS2 N2
	N2 T5 DE → C2 V2 CS2
C4	C2 N3 DE → V3 CS1 T4
	CS1 N3 DE → C2 V3 T4
	C2 T4 DE → V3 CS1 N3
	N3 T4 DE → C2 V3 CS1
C5	C3 N3 DE → V1 CS2 T6
	V1 CS2 N3 DE → C3 T6
	C3 T6 DE → V1 CS2 N3
	V1 T6 DE → C3 CS2 N3
	CS2 T6 DE → C3 V1 N3
	N3 T6 DE → C3 V1 CS2
C6	CS3 DE → C1 V1 N3 T5
	C1 V1 N3 T5 DE → CS3
C7	C1 V1 CS2 DE → N2 T4
	C1 N2 DE → V1 CS2 T4
	V1 N2 DE → C1 CS2 T4
	V1 T4 DE → C1 CS2 N2
	N2 T4 DE → C1 V1 CS2
C8	V2 N3 DE → C1 CS2 T5
	C1 CS2 N3 T5 DE → V2
C9	V1 N1 DE → C3 CS2 T5
	C3 T5 DE → V1 CS2 N1
	V1 CS2 T5 DE → C3 N1
	N1 T5 DE → C3 V1 CS2
C10	C1 N1 DE → V2 CS2 T4
C11	C3 V2 DE → CS1 N1 T4
	C3 CS1 DE → V2 N1 T4
	V2 CS1 DE → C3 N1 T4
	CS1 N1 DE → C3 V2 T4
	C3 T4 DE → V2 CS1 N1
C14	V3 DE → C2 CS1
	C2 CS1 DE → V3
C18	C1 N3 DE → T5
	C1 T5 DE → N3
	N3 T5 DE → C1
C21	V2 T5 DE → CS2
C23	C3 V1 DE → CS2
	C3 CS2 DE → V1
C24	C1 V2 DE → CS2
C25	C1 T4 DE → CS2
	CS2 T4 DE → C1
C27	V2 N1 DE → T4
	V2 T4 DE → N1
	N1 T4 DE → V2
C45	{} → DE

decision context. For this purpose, we have proposed an algorithm namely ECC algorithm which uses entropy-based precedence order of attributes. We have validated the reliability of the compressed concepts using percentage of error on deviance measures of intent weights. Our analysis found that the method is suitable to reduce the complexity of the concept lattice namely the number of concepts (nodes) with enough control on the consistency and the amount of information produced. We noted that the set of compressed concepts contains all the implications of the original concept lattice, and thus assures the lossless information in the output. As a future work, one can perform such compressions on different types of training datasets.

References

Assaghir, Z., M. Kaytoue, N. Messai, and A. Napoli. 2009. On the mining of numerical data with Formal Concept Analysis and similarity. *Société Francophone de Classification* 121–124.

Babin, M. A., and S. O. Kuznetsov. 2012. Approximating concept stability. Proceedings of the International Conference on Formal Concept Analysis, May, 7–15. Berlin/Heidelberg: Springer, Leuven, Belgium.

Belohlavek, R., and J. Macko. 2011. Selecting important concepts using weights. Proceedings of the International Conference on Formal Concept Analysis, May, 65–80. Berlin/Heidelberg: Springer, Nicosia, Cyprus.

Belohlavek, R., and V. Vychodil. 2009. Formal concept analysis with background knowledge: Attribute priorities. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*39 (4):399–409. doi:10.1109/TSMCC.2008.2012168.

Carpineto, C., and G. Romano. 2004. *Concept data analysis: Theory and applications.* John Wiley & Sons, Chichester, England.

Ch, A. K., S. M. Dias, and N. J. Vieira. 2015. Knowledge reduction in formal contexts using non-negative matrix factorization. *Mathematics and Computers in Simulation* 109:46–63. doi:10.1016/j.matcom.2014.08.004.

Cheung, S. K. 2004. Complexity reduction in lattice-based information retrieval: Theory, prototype development, and evaluation. Doctoral dissertation, City University of Hong Kong.

Csiszar, I. 2008. Axiomatic characterizations of information measures. *Entropy* 10 (3):261–73. doi:10.3390/e10030261.

Davey, B. A., and H. A. Priestley. 2002. *Introduction to lattices and order.* Cambridge University Press, Cambridge, UK.

Dias, S. M., and N. J. Vieira. 2013. Applying the JBOS reduction method for relevant knowledge extraction. *Expert Systems with Applications* 40 (5):1880–87. doi:10.1016/j.eswa.2012.10.010.

Fan, T. F., D. R. Liu, and G. H. Tzeng. 2007. Rough set-based logics for multicriteria decision analysis. *European Journal of Operational Research* 182 (1):340–55. doi:10.1016/j.ejor.2006.08.029.

Ganter, B., and R. Wille. 1989. Conceptual scaling. In *Applications of combinatorics and graph theory to the biological and social sciences*, 139–67. New York, US: Springer.

Ganter, B., and R. Wille. 1999. *Formal concept analysis: Mathematical foundations.* (translated from the German by Cornelia Franzke). Ed., Fred S. Roberts, New York, US: Springer-Verlag.

Hall, M. A., and G. Holmes. 2003. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, 15 (6):1437–47. doi:10.1109/TKDE.2003.1245283.

Han, J., and M. Kamber. 2006. *Data mining, southeast asia edition: Concepts and techniques.* New York, US: Morgan kaufmann.

Kumar, C. 2011. Knowledge discovery in data using formal concept analysis and random projections. *International Journal of Applied Mathematics and Computer Science* 21 (4):745–56. doi:10.2478/v10006-011-0059-1.

Kumar, C. A. 2012. Fuzzy clustering-based formal concept analysis for association rules mining. *Applied Artificial Intelligence* 26 (3):274–301. doi:10.1080/08839514.2012.648457.

Kumar, C. A., M. S. Ishwarya, and C. K. Loo. 2015. Formal concept analysis approach to cognitive functionalities of bidirectional associative memory. *Biologically Inspired Cognitive Architectures* 12:20–33. doi:10.1016/j.bica.2015.04.003.

Kumar, C. A., and P. K. Singh. 2014. Knowledge representation using formal concept analysis: A study on concept generation. In *Global trends in intelligent computing research and development*, Eds., B. K. Tripathi, and D. P. Acharya, IGI Global Press, Hershey, Pennsylvania, USA, 11, 306-336. DOI: 10.4018/978-1-4666-4936-1

Kumar, C. A., and S. Srinivas. 2010. Concept lattice reduction using fuzzy K-means clustering. *Expert Systems with Applications* 37 (3):2696–704. doi:10.1016/j.eswa.2009.09.026.

Kumar, C. A., and K. Sumangali. 2012. Performance evaluation of employees of an organization using formal concept analysis. Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), Salem, India, 94–98. IEEE, March.

Kuznetsov, S. O. 2001. On computing the size of a lattice and related decision problems. *Order* 18 (4):313–21. doi:10.1023/A:1013970520933.

Kuznetsov, S. O. 2007. On stability of a formal concept. *Annals of Mathematics and Artificial Intelligence* 49 (1–4):101–15. doi:10.1007/s10472-007-9053-6.

Kuznetsov, S. O., and S. A. Obiedkov. 2002. Comparing performance of algorithms for generating concept lattices. *Journal of Experimental and Theoretical Artificial Intelligence* 14 (2–3):189–216. doi:10.1080/09528130210164170.

Li, J., Z. He, and Q. Zhu. 2013. An entropy-based weighted concept lattice for merging multi-source geo-ontologies. *Entropy* 15 (6):2303–18. doi:10.3390/e15062303.

Li, J., C. A. Kumar, C. Mei, and X. Wang. 2017. Comparison of reduction in formal decision contexts. *International Journal of Approximate Reasoning* 80:100–22. doi:10.1016/j.ijar.2016.08.007.

Li, C., J. Li, and M. He. 2014. Concept lattice compression in incomplete contexts based on K-medoids clustering. *International Journal of Machine Learning and Cybernetics* 7 (4): 1–14. doi:10.1007/s13042-014-0288-3

Li, J., C. Mei, J. Wang, and X. Zhang. 2014. Rule-preserved object compression in formal decision contexts using concept lattices. *Knowledge-Based Systems* 71:435–45. doi:10.1016/j.knosys.2014.08.020.

Lv, Y. J., H. M. Liu, and J. H. Li. 2009. Attribute reduction of formal context based on concept lattice. Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery, vol. 1, Tianjin, China, 576–80. IEEE, August.

Messai, N., M. D. Devignes, A. Napoli, and M. Smail-Tabbone. 2008. Many-valued concept lattices for conceptual clustering and information retrieval. ECAI, vol. 178, Patras, Greece 127–31, July.

Oruç, O. E., E. Kuruoğlu, and O. Vupa. 2009. An application of entropy in survey scale. *Entropy* 11 (4):598–605. doi:10.3390/e11040598.

Priss, U. 2006. Formal concept analysis in information science. *Arist* 40 (1):521–43.

Rice, M. D., and M. Siff. 2001. Clusters, concepts, and pseudometrics. *Electronic Notes in Theoretical Computer Science* 40:323–46. doi:10.1016/S1571-0661(05)80060-X.

Rifkin, J. 1980. *Entropy: A new world view.* New York, US: Viking Press

Rooker, J. C., and D. A. Gorard. 2007. Errors of intravenous fluid infusion rates in medical inpatients. *Clinical Medicine* 7 (5):482–85. doi:10.7861/clinmedicine.7-5-482.

Roth, C., S. Obiedkov, and D. Kourie. 2008. Towards concise representation for taxonomies of epistemic communities. In *Concept lattices and their applications*, Eds., R. Belohlavek, S. O. Kuznetsov, 240–55. Olomouc, Czech Republic, Europe, Berlin/Heidelberg:Springer.

Shannon, C. E. 1948. A mathematical theory of communication, bell System technical Journal 27: 379–423 and 623–656. *Mathematical Reviews (Mathscinet): MR10*133e.

Shivhare, R., and A. K. Cherukuri. 2017. Three-way conceptual approach for cognitive memory functionalities. *International Journal of Machine Learning and Cybernetics* 8(1), 21–34. doi:10.1007/s13042-016-0593-0.

Singh, P. K., C. Aswani Kumar, and A. Gani. 2016. A comprehensive survey on formal concept analysis, its research trends and applications. *International Journal of Applied Mathematics and Computer Science* 26 (2):495–516. doi:10.1515/amcs-2016-0035.

Singh, P. K., A. K. Cherukuri, and J. Li. 2015. Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy. *International Journal of Machine Learning and Cybernetics* 6(1): 1–11. doi: 10.1007/s13042-014-0313-6

Singh, P. K., and C. A. Kumar. 2016. Concept lattice reduction using different subset of attributes as information granules. In *Granular computing*, 1–15. doi:10.1007/s41066-016-0036-z

Soldano, H., V. Ventos, M. Champesme, and D. Forge. 2010. Incremental construction of alpha lattices and association rules. Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 351–60). Springer Berlin Heidelberg, Cardiff, UK, September, 351–360.

Stumme, G. 2002. Efficient data mining based on formal concept analysis. Proceedings of the International Conference on Database and Expert Systems Applications, 534–46. Springer Berlin Heidelberg, Aix-en-Provence, France, September.

Sumangali, K., and C. A. Kumar. 2013. Critical Analysis on Open Source LMSs using FCA. *International Journal of Distance Education Technologies (IJDET)*11 (4):97–111. doi:10.4018/IJDET.

Sumangali, K., and C. Kumar. 2014. Determination of interesting rules in FCA using information gain. Proceedings of the 1st International Conference on Networks and Soft Computing (ICNSC), Guntur, India 304–08. IEEE, August.

Tripathy, B. K., D. P. Acharjya, and V. Cynthya. 2013. A framework for intelligent medical diagnosis using rough set with formal concept analysis. International Journal of Artificial Intelligence and Applications 2(2): 45–66.

Wille, R. 1982. Restructuring lattice theory: An approach based on hierarchies of concepts. In *Ordered sets*, 445–70. Netherlands: Springer.

Zhang, S., P. Guo, J. Zhang, X. Wang, and W. Pedrycz. 2012. A completeness analysis of frequent weighted concept lattices and their algebraic properties. *Data and Knowledge Engineering*, Ed., I. Rival, Reidel, 81:104–17. doi:10.1016/j.datak.2012.08.002.

Zhang, S., and X. Wu. 2011. Fundamentals of association rules in data mining and knowledge discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (2):97–116.